

GUÍA
básica de la

IA

Primera edición

GUÍA BÁSICA DE LA IA

© de los textos, Aguilar, I.; Alepuz, V.; Alfaro, J.; Bañón, J.J.; Botti, V.; Despujol, I.; Giménez, J.; Linares, J.; Linares, J.M.; Majadas, V.; Martínez, J.; Monsoriu, M.; Montesa, E.; Morillas, C.; Muñoz, J.M.; Ortega, J.; Ortuño, A.; Peñarrubia, J.P.; Plasencia, A.; Rieta, J.; Sales, M.; Segarra, R. (2024).

Edita Smart Digital. Disponible en: <https://www.smartdigital.es/>

Imprime: 315 gr laboratorio gráfico SL

ISBN: 978-84-19814-42-5

Primera edición de diciembre 2024

Cómo citar un epígrafe específico. Ejemplo: Peñarrubia, J.P. (2024). *Ética de los sistemas de IA: Construyendo una IA transparente, equitativa y fiable*. En Smart Digital (Ed.) (2024) *Guía básica de la IA*. pp.249-260. Disponible en: <https://www.smartdigital.es/>

Actualizaciones de la bibliografía y casos de éxito en [smartdigital.es](https://www.smartdigital.es)

Nota de responsabilidad: El texto y los contenidos gráficos incluidos en cada epígrafe específico de esta publicación son responsabilidad de su correspondiente autor o autores.

No se permite la reproducción total o parcial de este libro, ni su incorporación a un sistema informático, ni su transmisión en cualquier forma o por cualquier medio, sea este electrónico, mecánico, por fotocopia, por grabación u otros métodos, sin el permiso previo y por escrito del editor. La infracción de los derechos mencionados puede ser constitutiva de delito contra la propiedad intelectual (Arts. 270 y siguientes del Código Penal). Las solicitudes para la obtención de dicha autorización total o parcial deben dirigirse a CEDRO (Centro Español de Derechos Reprográficos).

GUÍA básica de IA



PRÓLOGO

🗨 Andrés Pedreño Muñoz

Nos encontramos en un momento crucial para el desarrollo de nuestras sociedades: la irrupción de la Inteligencia Artificial (IA) como motor transformador no sólo redefine el paradigma tecnológico, sino que también establece nuevas bases para el crecimiento económico y la competitividad global. La IA promete impulsar la productividad y generar valor en proporciones extraordinarias. Estudios recientes de McKinsey¹ proyectan que la IA generativa podría sumar entre 2.6 y 4.4 billones de dólares anuales en valor agregado en más de 60 áreas de aplicación, duplicando esta cifra cuando se considera el impacto en el sector del software. Este poder disruptivo no es solo una promesa; es ya una realidad que debemos comprender y gestionar si queremos estar a la vanguardia.

La IA nos ofrece una herramienta única para abordar algunos de los desafíos más urgentes de la humanidad. Desde la mitigación de los efectos del cambio climático hasta la regeneración de ecosistemas y la gestión eficiente de recursos críticos como el agua y la energía, la IA es clave en la búsqueda de soluciones sostenibles. En el ámbito de la salud, la IA abre nuevas posibilidades en la lucha contra enfermedades complejas y permite avanzar hacia una medicina personalizada que responde a las necesidades individuales de cada paciente. Además, en los sectores productivos tradicionales, desde el turismo hasta la banca, la IA redefine los estándares de competitividad, permitiendo a las empresas que adopten estas tecnologías liderar en sus mercados y adaptarse a las demandas de un entorno global cada vez más digital.

Europa, sin embargo, enfrenta un reto en su posicionamiento en esta nueva *liga de campeones* tecnológica, frente a gigantes como Estados Unidos y China. Mientras estos dos países han hecho apuestas decididas y multimillonarias para liderar en IA, Europa ha mostrado un enfoque marcado por la

1 <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year>

regulación y la protección de derechos fundamentales, lo cual es crucial pero insuficiente si deseamos cerrar la brecha competitiva. Nos falta ambición presupuestaria, padecemos una limitada inversión en I+D que nos ha relegado a un segundo plano. Pero sobre todo faltan empresas relevantes, no tenemos gigantes tecnológicos, apenas unicornios y nuestras startups se enfrentan a innumerables desventajas a la hora de escalar y crecer.

Sin embargo, esta situación también ofrece una oportunidad estratégica: España y, en particular, la Comunidad Valenciana tienen la posibilidad de posicionarse como referentes en el ámbito de la IA europea, promoviendo un ecosistema en el que la innovación y la ética coexistan y se potencien mutuamente, tal como sugiere el reciente *Informe Draghi*. Este informe reivindica acertadamente la necesidad de fomentar inversiones en torno a los 800.000 millones de euros anuales para superar el gap de Europa con China y Estados Unidos.

En este contexto, la *Guía Básica de la IA* cobra un valor fundamental. La colaboración de los colegios profesionales en su elaboración demuestra un compromiso encomiable con la capacitación y el progreso de nuestra sociedad. Esta guía no solo acerca a los profesionales de diversas áreas a los conceptos clave de la IA, sino que también incorpora una valiosa colección de casos de éxito que ilustran cómo se están aplicando soluciones de IA en sectores como la salud, la educación, el comercio y la industria. Estos ejemplos representan no solo un testimonio del potencial de la IA en el ámbito práctico, sino también un incentivo para que más organizaciones se sumen a esta transformación digital, sirviendo como un efecto demostración en nuestro tejido empresarial y social.

La interdisciplinariedad que caracteriza a esta guía, apoyada en el trabajo conjunto de profesionales de ámbitos diversos, enriquece su contenido y ofrece una visión holística que es crucial en una tecnología tan compleja y multifacética como la IA. Los colegios profesionales han aportado sus conocimientos y experiencia, permitiendo que esta guía se nutra de una pluralidad de perspectivas que refuerzan su utilidad práctica. Este esfuerzo conjunto no solo facilita una comprensión accesible de la IA, sino que también fomenta un espacio para la colaboración y el intercambio de conocimiento, elementos esenciales para el progreso tecnológico y social de nuestra comunidad.

Estoy convencido de que esta obra será un recurso esencial para los profesionales que deseen comprender y aplicar la IA en sus respectivos campos. Más aún, espero que inspire a otros a explorar, adoptar y adaptar la IA, con el propósito de fortalecer nuestro tejido productivo y promover un crecimiento que, en lugar de responder a modelos de competencia desmesurada, integre valores de sostenibilidad, ética e inclusión. Esta guía es, en definitiva, un paso firme hacia un futuro donde la IA sea un motor de innovación responsable y de progreso para todos. 📄

INTRODUCCIÓN

PROMOVER EL FUTURO DIGITAL DE LA COMUNIDAD VALENCIANA

En la Era Digital, la Inteligencia Artificial (IA) se ha convertido en una herramienta esencial que transforma todos los aspectos de nuestra vida cotidiana y profesional. Para asegurar que la Comunidad Valenciana (CV) no se quede atrás en esta revolución tecnológica, quince colegios profesionales han unido fuerzas a través del Acuerdo Marco SmartDigital. Este acuerdo tiene un objetivo claro: promover la transformación digital y fortalecer las competencias digitales de los profesionales colegiados en la CV. Como parte de esta colaboración, se ha decidido elaborar esta **Guía Básica de la IA**, un recurso diseñado para ofrecer una comprensión accesible y completa de los conceptos fundamentales de la IA y sus aplicaciones en diversos sectores.

La *Guía Básica de la IA* nace de la necesidad de proporcionar una base sólida de conocimientos sobre la Inteligencia Artificial a aquellos profesionales que desean entender cómo esta tecnología puede aplicarse en sus respectivos campos. A medida que la IA se integra cada vez más en los procesos empresariales, administrativos y educativos, es crucial que los profesionales estén bien informados sobre sus capacidades, oportunidades y desafíos.

Los objetivos principales de la guía son tres:

1. ***Difusión de conceptos clave:*** La IA es un campo amplio y en constante evolución. Para quienes no están familiarizados con ella, puede resultar abrumadora. La guía desglosa los conceptos fundamentales de manera sencilla y clara, abarcando temas como el aprendizaje automático, el aprendizaje profundo, el Big Data y los algoritmos de IA. Al simplificar estos términos, se facilita a los profesionales una mejor comprensión de cómo funciona la IA y cómo puede aplicarse en su área de especialización.
2. ***Puesta en común de conocimientos:*** La guía también funciona como un punto de encuentro para compartir conocimientos y experiencias entre diferentes colegios profesionales. Incluye casos de éxito y aplicaciones

prácticas de la IA en sectores como la salud, la educación, el comercio y la industria. Estos ejemplos muestran cómo la IA está siendo utilizada para resolver problemas específicos y mejorar la eficiencia y efectividad en diversos ámbitos, inspirando a los profesionales a considerar soluciones similares en sus propios campos.

3. **Fomento de la Transformación Digital:** Un objetivo fundamental de la guía es impulsar la transformación digital en la Comunidad Valenciana. Al ofrecer una comprensión accesible de la IA y sus aplicaciones, se pretende dotar a los profesionales con las herramientas necesarias para adoptar y aprovechar esta tecnología en su trabajo diario. La guía incluye recomendaciones y estrategias prácticas para integrar la IA de manera efectiva, ayudando así a las organizaciones a mantenerse competitivas en un entorno cada vez más digital.

La creación de esta *Guía Básica de la IA* ha sido posible gracias a la colaboración desinteresada de un numeroso grupo de profesionales cuyos nombres constan en los créditos del libro y datos de filiación en la lista que aparece al final de la publicación (página 466). Hay que señalar que aunque la mayor parte de los autores son miembros de los colegios que forman parte del *Acuerdo Marco Intercolegial Smart Digital* (ver listado alfabético en página 470), también han colaborado destacados expertos de otras entidades afines. La *Guía* ha ganado así en profundidad y visión plural, fortaleciendo sus contenidos y reflejando el espíritu de cooperación en el que se sustenta este proyecto.

Deseamos manifestar de forma pública nuestro reconocimiento a los siguientes profesionales que aún sin pertenecer a los colegios de Smart Digital¹, creyeron desde el primer momento en el proyecto:

- Mar Monsoriu, periodista y miembro de la Asociación de Periodistas y Profesionales de la Comunicación Valencianos (APPV)²

1 <https://smartdigital.es/>

2 <https://www.appperiodistasvalencianos.es/>

- José Manuel Muñoz, abogado y miembro de Adequa³
- Adolfo Plasencia, periodista científico y autor de *De neuronas a galaxias*⁴;
- Vicent Botti, catedrático y director del Instituto Universitario Valenciano de Investigación en Inteligencia Artificial (VRain⁵) ;
- Jordi Linares, Doctor en Informática y profesor de la Universitat Politècnica de València (Campus de Alcoy)
- Manuel Sales, presidente de AvalBit⁶ e ingeniero de telecomunicaciones.

Nuestro agradecimiento a Andrés Pedreño⁷, quien tan pronto conoció el proyecto se brindó a escribir el prólogo de la Guía.

Gracias a los conocimientos y visiones de todos los autores, esta guía ofrece una perspectiva amplia y especializada que esperamos sea de gran utilidad para los profesionales de la Comunidad Valenciana y más allá.

3 <https://adequa.eu/equipo/jose-manuel-munoz-vela/>

4 <https://deneuronasagalaxias.com/>

5 <https://vrain.upv.es/>

6 <https://avalbit.org/>

7 <https://ost.torrejuana.es/>

Estructura de la Guía

La *Guía Básica de la IA* está estructurada para proporcionar una comprensión integral del campo:

CAP	TITULO	DETALLE
I	Inicios y evolución de la IA	Conceptos y evolución
II	Tipos de IA	Se describen las diferentes IA habitualmente mencionadas en la: débil (ANI), general (AGI) y súper (ASI).
III	Fundamentos de la IA	Se cubren aspectos esenciales relacionados con las metodologías de IA, los entornos y herramientas utilizados, así como los casos de uso en el mundo real
IV	Desafíos	Se exploran las limitaciones y problemas actuales que enfrenta la IA, incluyendo cuestiones técnicas y sociales.
V	Ética y Regulación	Se analiza la importancia de la ética en el desarrollo de la IA y las regulaciones necesarias para asegurar su uso responsable.
VI	Futuro	Se discuten las posibles direcciones y el impacto futuro de la IA en la sociedad.
VII	Quiero usar la IA	Breve introducción a los principales modelos de IA generativa con recomendaciones prácticas y recursos formativos.
VIII	Casos de Éxito	Una colección de casos seleccionados por los colegios que participan en la Guía

Cuadro nº1: Visión sinóptica de la Guía

La Guía también incluye como anexos, un glosario de conceptos y términos imprescindibles, seguido de una bibliografía con los textos de referencia, así como un listado de los autores que han contribuido con los textos y casos que contiene la Guía. 

00. TABLA DE CONTENIDOS

015	01. INICIOS Y EVOLUCIÓN DE LA IA
031	02. TIPOS DE INTELIGENCIA ARTIFICIAL
043	03. FUNDAMENTOS DE LA IA
045	Macrodatos (Big Data)
057	Redes neuronales Artificiales
067	Aprendizaje automático (ML: Machine learning)
087	El Deep Learning en Sistemas de CCTV
105	Procesamiento de Lenguaje natural (PLN)
111	Grandes modelos de lenguaje (LLM)
121	Visión por computador (CV: Computer Vision)
131	Los robots y la IA
137	Sistemas expertos
149	Agentes Inteligentes y los Sistemas Multiagente
155	IA en Ciberseguridad
169	04. DESAFÍOS DE LA IA
171	Desafíos e impacto de la IA
181	Impacto en los sectores económicos clave
193	La IA en la comunicación: adoptar lo mejor de la IA y gestionar sus riesgos
201	Impacto en la sociedad
207	Inteligencia Artificial Generativa en la Educación Superior
217	Auditoría, contabilidad y asesoramiento fiscal y financiero
233	IA y geopolítica
247	05. ÉTICA Y REGULACIÓN DE LA IA
249	Ética de los sistemas de IA: Construyendo una IA transparente, equitativa y fiable
261	Impacto de la IA en la sociedad: empleo, administración, ciudadanía y política
271	Regulación de la IA
283	Gobernanza de la IA

291	06. FUTURO DE LA IA
293	Impacto de la IA en la sociedad del futuro
309	Inteligencia Alternativa (IA): El último paso de la Humanidad
317	IA: avances recientes y tendencias
325	La Era de la Inteligencia Artificial (IA) y el futuro de la Humanidad
331	Educación e Inteligencia Artificial: Transformando la Sociedad del Futuro
341	Usar la IA con inteligencia
349	Aproximación crítica a la IA Generativa y sus sucedáneos
401	07. QUIERO TRABAJAR CON LA IA. ¿POR DÓNDE EMPIEZO?
415	08. CASOS DE ÉXITO
417	Proyecto Khanmigo
421	Proyecto ZG3D
425	Aplicación práctica de la IA en el ámbito de la Economía Circular
434	Avatares web conectados a chatbot
437	Integración de chatgpt
440	Avatares Personalizados para Videos Educativos y Publicitarios
443	Detección automática de cáncer de mama
447	Proyecto BISOC
450	Proyecto LIDAR
453	Onboarding Digital
457	09. ANEXOS
459	Glosario imprescindible de la IA
466	Lista de autores de la Guía Básica de la IA
470	Listado alfabético de los Colegios Profesionales que integran Smart Digital
472	Bibliografía

INICIOS Y EVOLUCIÓN DE LA INTELIGENCIA ARTIFICIAL

🗨 César Morillas, José Ortega y Enric Montesa

EVENTO FUNDACIONAL DE LA IA

El nacimiento de la *inteligencia artificial* (IA) está marcado por un evento clave: el *Encuentro de Dartmouth*, celebrado en el verano de 1956 en el Dartmouth College de New Hampshire. Este encuentro es considerado el punto de partida formal de la disciplina de la IA. Allí se reunieron los diez visionarios que concibieron las primeras ideas sobre la posibilidad de crear máquinas capaces de “pensar”. Los participantes de esta histórica conferencia (matemáticos, ingenieros, psicólogos, politólogos y economistas) abordaron por primera vez los fundamentos que hoy conocemos como *inteligencia artificial*.

CONTEXTO DE LA CONFERENCIA DE DARTMOUTH

El evento fue impulsado principalmente por el matemático *John McCarthy*, quien acuñó el término *inteligencia artificial*, juntamente con otros pioneros como *Marvin Minsky*, *Claude Shannon* y *Nathaniel Rochester*. La propuesta central de McCarthy era que *“cualquier aspecto del aprendizaje o cualquier otra característica de la inteligencia puede, en principio, ser descrito con tal precisión que una máquina puede ser construida para simularlo”*. Esta ambiciosa premisa fue el motor del debate durante el encuentro y, aunque no se lograron avances tecnológicos inmediatos, las bases teóricas sentadas allí marcaron un antes y un después en el campo de la computación y la inteligencia artificial.

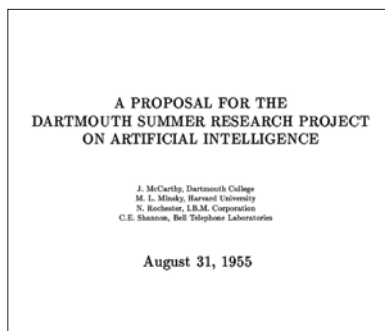
Dartmouth College, Hanover, New Hampshire (EEUU)



EL "PROYECTO"¹

El documento preparado por John McCarthy titulado "*Una Propuesta para el Proyecto de Investigación de Verano de Dartmouth sobre Inteligencia Artificial*"² es un texto fundamental en la historia de la inteligencia artificial (IA). En él, se destacaban los siguientes puntos clave:

- *Objetivo del proyecto:* investigar la posibilidad de que las máquinas puedan simular cualquier aspecto de la inteligencia humana. McCarthy y sus colegas creían que la inteligencia artificial podía ser abordada mediante la creación de programas que imitaran procesos cognitivos humanos.
- *Propuesta de investigación:* la inteligencia artificial podía ser abordada mediante una *combinación de programación y experimentación*, sugiriéndose que se podía avanzar significativamente en el entendimiento de la IA si se reunían los recursos y expertos adecuados durante un encuentro de verano.
- *Importancia y expectativas:* McCarthy argumentaba que, si el proyecto tenía éxito, podía llevar a avances importantes en diversas áreas de la ciencia y la tecnología. El éxito del proyecto tendría un *impacto profundo en la manera en que se entenderían e iban a desarrollarse las máquinas inteligentes*.
- *Financiación:* Se especificaban los recursos económicos necesarios para llevar adelante el proyecto, sugiriéndose el patrocinio de la *Fundación Rockefeller* y la colaboración de *Bell Labs* e *IBM* (empresas en las que trabajaban dos de los promotores del encuentro: Claude Shannon y Nathaniel Rochester).



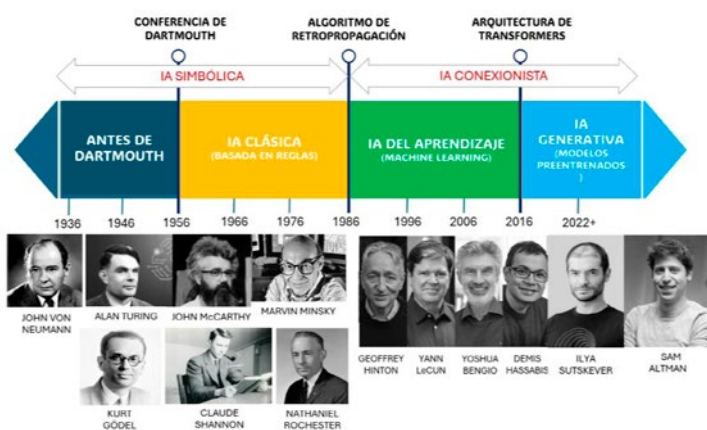
1 McCarthy et alia (1955): *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. Disponible en: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>

2 PDSRPAI es el acrónimo ocasionalmente usado para referirse a la propuesta: "*Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*".

EVOLUCIÓN DE LA IA

Una forma muy sencilla de abordar la historia de la IA es hablando de sus dos etapas, cuya divisoria queda establecida en 1986, año en el que *Geoffrey Hinton*³ junto a *Rumelhart* y *Williams*, publicó su famoso y crucial *paper*⁴ en el avance del *Aprendizaje Profundo* (Deep Learning), y en el que introdujo la retropropagación (*backpropagation*) como un método efectivo para entrenar redes neuronales artificiales mediante la minimización de errores a través del ajuste de los pesos⁵ en las capas de la red.

LINEA DE TIEMPO DE LA INTELIGENCIA ARTIFICIAL



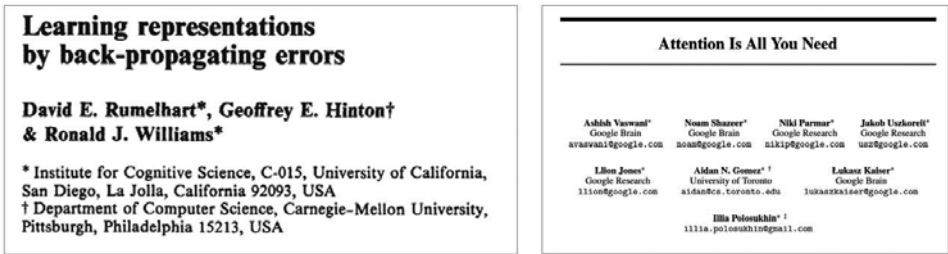
La rapidez de los avances en inteligencia artificial generativa, basada en *modelos pre-entrenados*, experimentó un gran impulso a partir de 2017 con el desarrollo de los *transformers*, una arquitectura clave en el campo de la IA.

3 Geoffrey Hinton, nacido el 6 de diciembre de 1947 en Wimbledon, Inglaterra, es conocido como el "padrino de la inteligencia artificial" por su trabajo pionero en redes neuronales y aprendizaje profundo. En 2024 ha recibido el Premio Nobel de Física por sus contribuciones a este campo.

4 Rumelhart, D., Hinton, G. & Williams, R. *Learning representations by back-propagating errors*. *Nature* 323, 533–536 (1986). <https://doi.org/10.1038/323533a0>

5 Un número, denominado peso, representa las conexiones entre un nodo y otro. El peso es un número positivo si un nodo estimula a otro, o negativo si un nodo suprime a otro. Los nodos con valores de peso más altos tienen mayor influencia en los demás nodos.

Ashish Vaswani junto a otros seis colegas, publicó el *paper*⁶ que dio a conocer el concepto de los *transformers* y el uso del mecanismo de atención, abriendo las puertas a un entrenamiento más eficiente de modelos pre-entrenados en grandes cantidades de datos.



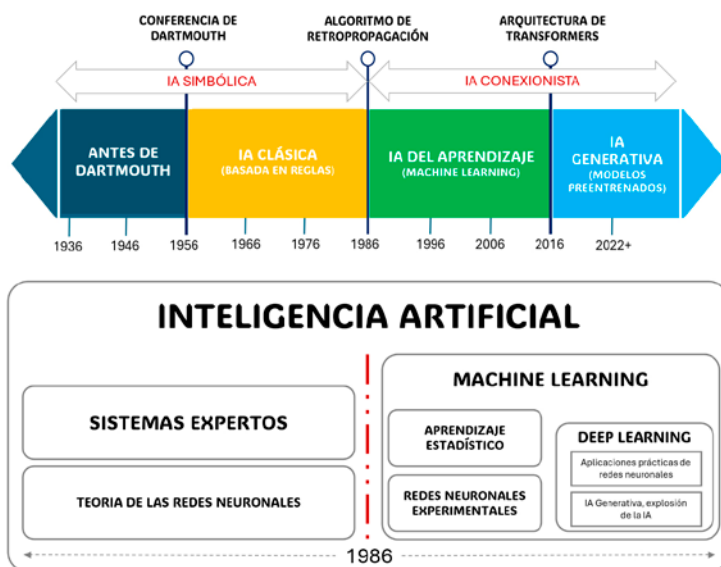
TÉCNICAS CLAVE DE LA IA CLÁSICA Y LA IA MODERNA

La metáfora del “cerebro electrónico” y de las “redes neuronales” tiene sus raíces en las ideas sobre la capacidad de las máquinas para razonar, plantea- das por *Alan Turing* en su influyente artículo de 1950⁷, “*Computing Machinery and Intelligence*”, y en el modelo de redes neuronales propuesto por *Warren McCulloch* y *Walter Pitts* en 1943⁸. Esta metáfora surgió antes de la Conferencia de Dartmouth de 1956 y puede haber inspirado a figuras clave como *John McCarthy*, uno de los pioneros en la IA. Sin embargo, en la primera etapa del desarrollo de la inteligencia artificial, conocida como **IA simbólica o basada en reglas**, los enfoques neuronales quedaron relegados, priorizándose los mo- delos formales y lógicos para emular la inteligencia.

Esta etapa temprana de la IA produjo avances notables, como la **teoría de juegos** (ajedrez y damas), los primeros **chatbots** (como Eliza), **resolvedo- res de problemas** (Logic Theorist), así como los lenguajes de programación LISP y PROLOG, que facilitaron el desarrollo de sistemas más complejos. Más

6 Ashish Vaswani et alia (2017): *Attention Is All You Need*. <https://arxiv.org/abs/1706.03762>
7 Turing, Alan (2012): *¿Puede pensar una máquina?* Oviedo. KRK Ediciones.
8 McCulloch, W.S., Pitts, W. *A logical calculus of the ideas immanent in nervous activity*. Bu- lletin of Mathematical Biophysics 5, 115–133 (1943). <https://doi.org/10.1007/BF02478259>

adelante, estos enfoques dieron lugar a los **sistemas expertos**, capaces de simular el razonamiento humano en dominios específicos, como la medicina y la química, marcando el auge de la IA en las décadas siguientes.



Gráficamente se puede ver a la IA como una caja en la hay numerosas aplicaciones ubicadas antes y después de 1986⁹:

- La primera etapa es conocida como la **IA Clásica o simbólica** y en ella se ubican los *Sistemas Expertos* y la teoría de las redes neuronales.
- La segunda etapa se conoce como **IA Moderna o conexionista**, y en ella aparece el *Aprendizaje Automático* (Machine Learning) con sus subcompartimentos del *Aprendizaje Estadístico*, las *Redes Neuronales* experimentales y el *Aprendizaje Profundo* (Deep Learning).

9 Flores, Antonio (2024): *Una mente infinita, La revolución de la Inteligencia Artificial*. Barcelona. Editorial Tusquets. Ver capítulo segundo y gráfico en página 26

A. Técnicas de la IA clásica

Este paradigma de la IA se centra en la manipulación de símbolos y la lógica formal para la toma de decisiones y resolución de problemas. Algunas de las técnicas principales son:

- **Sistemas de reglas:** Los sistemas expertos y los sistemas basados en reglas utilizan una base de conocimiento compuesta por hechos y reglas predefinidas (del tipo "si-entonces") para tomar decisiones. Un ejemplo famoso fue MYCIN, un sistema experto para el diagnóstico de infecciones bacterianas.
- **Lógica proposicional y de predicados:** se basa en el uso de la lógica formal para representar y razonar sobre el conocimiento. Permite hacer inferencias lógicas. Los "*Resolvedores de Lógica*" y las "*Pruebas Automáticas de Teoremas*" son herramientas potentes para automatizar el razonamiento lógico y la verificación formal. Aunque son más complejos y menos intuitivos que las técnicas modernas de IA como el *Aprendizaje Automático*, siguen siendo fundamentales en áreas donde la precisión lógica es esencial.
- **Búsqueda en espacios de estados:** se trata de los algoritmos que buscan soluciones en grandes espacios de posibles estados o configuraciones. Técnicas como búsqueda en profundidad y amplitud fueron ampliamente utilizadas. Ejemplos: Problemas de juego (ajedrez), rompecabezas de 8 piezas.
- **Redes Bayesianas:** son redes probabilísticas que permiten inferencias en dominios de incertidumbre. Esta técnica permite modelar el conocimiento de forma probabilística. Ejemplo: Diagnóstico médico basado en probabilidades.
- **Algoritmos de búsqueda heurística:** Algoritmos como A^* usan heurísticas para encontrar soluciones óptimas de manera más eficiente. Ejemplo: Resolución de problemas de rutas.

B. Técnicas de la IA Moderna

Se basa en el uso de grandes cantidades de datos, *Aprendizaje Automático* (Machine Learning) y técnicas de procesamiento computacional avanzado.

Los avances en hardware y la disponibilidad de datos han permitido que la IA moderna, particularmente el *Aprendizaje Profundo* (Deep Learning), se desarrolle rápidamente.

- **Aprendizaje Automático** (Machine Learning): Los algoritmos que aprenden patrones a partir de datos y hacen predicciones sin ser programados explícitamente. Métodos como regresión lineal, árboles de decisión y máquinas de soporte vectorial son comunes. Ejemplos: Clasificadores de imágenes, predicción de ventas.
- **Aprendizaje Profundo** (Deep Learning): Es un subcampo del aprendizaje automático que utiliza redes neuronales profundas con múltiples capas para extraer características jerárquicas de los datos. Las redes neuronales convolucionales (CNN) y las redes neuronales recurrentes (RNN) son ejemplos de este enfoque. Ejemplo: Reconocimiento de voz, clasificación de imágenes, procesamiento de lenguaje natural (PLN).
- **Aprendizaje por Refuerzo** (Reinforcement Learning): se basa en la capacidad de un agente que aprende a tomar decisiones secuenciales en un entorno, recibiendo recompensas o penalizaciones. Se utiliza en juegos, robótica y automatización. - Ejemplos: AlphaGo, sistemas de control autónomo.
- **Modelos Generativos**: consisten en un tipo de algoritmos que pueden aprender a generar datos similares a los que se entrenaron. Las redes generativas adversarias (GANs) y los modelos *autoregresivos* (como GPT) son los más destacados del presente. Ejemplo: Generación de imágenes, texto, audio.
- **Procesamiento de Lenguaje Natural (PLN)**: Se trata de un subcampo de la IA que se enfoca en la comprensión y generación del lenguaje humano. Los modelos modernos, como los transformadores (p. ej., **BERT**, **GPT**), han revolucionado este campo. Ejemplos: *Chatbots*, traducción automática, análisis de sentimientos.
- **Transferencia de Aprendizaje**: es una técnica donde un modelo entrenado en una tarea se reutiliza (parcial o completamente) para una nueva tarea, mejorando la eficiencia en problemas con datos limitados. Ejemplos: Modelos *preentrenados* como **BERT** o **GPT** aplicados a distintas tareas de **PLN**.

Se puede concluir señalando que mientras la *IA Clásica* se basa en reglas explícitas y lógica formal, la *IA Moderna* utiliza modelos de datos masivos que aprenden automáticamente. Pero es una divisoria que no debe cegarnos y hacernos olvidar las virtudes de la *IA Clásica* en el ámbito de los problemas estructurados con reglas claras. No obstante, el triunfo de la *IA Moderna* (conexionista) sobre la *IA Clásica* (simbólica) estriba en su mayor eficacia en la resolución de problemas complejos que requieren manejo de datos no estructurados, como imágenes o lenguaje.

CUATRO COMPONENTES CLAVE DE LA IA

1. Computación

Es muy conocida la anécdota de *Alan Turing* ejecutando de cabeza el algoritmo que había inventado para jugar al ajedrez, debido a inexistencia de máquinas capaces de poderlo ejecutar.

La evolución de la computación ha avanzado enormemente desde los primeros procesadores hasta las tecnologías especializadas que tenemos hoy en día. Al principio, las **CPU** (unidades centrales de procesamiento) eran el núcleo de los sistemas informáticos. Estas se encargaban de realizar todas las tareas de cálculo general. Con el tiempo, surgieron las **GPU** (unidades de procesamiento gráfico), diseñadas originalmente para mejorar el rendimiento en gráficos y videojuegos, acabaron demostrando finalmente que también eran muy eficientes en el procesamiento de grandes cantidades de datos en paralelo, algo ideal para entrenar modelos de inteligencia artificial.

En la actualidad, se ha dado un paso más allá con el desarrollo de hardware especializado como las **NPU** (unidades de procesamiento neuronal) y las **TPU** (unidades de procesamiento tensorial), que están diseñadas específicamente para ejecutar algoritmos de inteligencia artificial y aprendizaje automático. A diferencia de las **CPU** y **GPU**, estas nuevas unidades están optimizadas para manejar las operaciones matemáticas complejas que los modelos de IA requieren, como las redes neuronales.

Lo más interesante es que muchos de estos algoritmos de IA pueden ser *embebidos* directamente en el hardware. Esto significa que los cálculos que antes requerían software especializado ahora se realizan directamente en los

circuitos, lo que hace que el procesamiento sea mucho más rápido y eficiente. Por ejemplo, un teléfono inteligente moderno puede tener una **NPU** que acelera las tareas de reconocimiento facial o la mejora de imágenes, sin depender completamente del software, lo que ahorra energía y tiempo.

Este avance ha permitido que la inteligencia artificial sea más accesible y rápida, impulsando desarrollos en áreas como los asistentes virtuales, los vehículos autónomos y el análisis de datos a gran escala.

2. Datos

La evolución de la IA pasó de sistemas basados en *conocimiento explícito* a modelos que aprenden a partir de *datos*. Los primeros enfoques, como los *Sistemas Expertos*, usaban reglas programadas a mano para imitar el razonamiento humano en áreas específicas. Aunque útiles, estos sistemas eran difíciles de escalar y limitados a dominios cerrados (micromundos).

Con el auge del *Aprendizaje Automático* y la disponibilidad de grandes volúmenes de datos (conocidos como macrodatos o Big Data), la IA dio un salto hacia *modelos estadísticos* que aprenden patrones directamente. Las técnicas modernas, como las *redes neuronales profundas*, superaron las limitaciones de los *Sistemas Expertos*, permitiendo que la IA se adaptase y mejorase con más datos, revolucionando de este modo las tareas complejas tales como el reconocimiento de imágenes y el procesamiento de lenguaje natural.

3. Algoritmos

Los primeros algoritmos de inteligencia artificial (reglas lógicas y búsqueda heurística) evolucionaron hacia los *Sistemas Expertos* que se hicieron muy populares en la década de los años 80's del siglo XX (codificación del conocimiento de especialistas humanos). Pero este enfoque se enfrentó al desafío de la "adquisición del conocimiento": extraer y codificar el conocimiento de los expertos humanos resultaba ser un proceso lento, costoso y propenso a errores.

Además, la rigidez del enfoque de los *Sistemas Expertos* frenaba su adopción generalizada ya que era incapaz de adaptarse a situaciones nuevas o

imprevistas, puesto que solo podía operar dentro de los límites estrictos de las reglas programadas.

Si a lo anterior se le sumamos los problemas de mantenimiento y la complejidad de sus constantes actualizaciones y revisiones por parte de expertos en el dominio (medicina, finanzas, etc.) y programadores, se comprende cómo la *IA Clásica* perdió su primogenitura y se vio obligada a entregar el testigo a la *IA Moderna* basada en el *Aprendizaje Automático* que tomó fuerza en los 90 y 2000, con técnicas como las *Redes Neuronales*, los *Árboles de Decisión* y las *Máquinas de Vectores de Soporte*, métodos que permitían a los algoritmos aprender patrones a partir de datos.

4. Grandes Modelos (Preentrenados)

La IA experimentó un gran salto adelante en la década de 2010 gracias al auge del *Aprendizaje Profundo*. Las *Redes Neuronales Profundas*, entrenadas con enormes cantidades de datos, lograron avances notables en tareas como la *Visión por Ordenador* (CV) y el *Procesamiento del Lenguaje Natural* (PLN).

Esto llevó al desarrollo de los modelos preentrenados actuales, como los *Transformers* que representan un cambio de paradigma en la IA. Se trata de *Redes Neuronales* masivas entrenadas en vastos conjuntos de datos no etiquetados, lo que les permite capturar patrones y relaciones complejas del lenguaje o las imágenes de manera general.

La clave de los modelos preentrenados es su capacidad de transferencia de aprendizaje. Después de su entrenamiento inicial, pueden ajustarse finalmente para tareas específicas con relativamente pocos datos adicionales. Esto ha democratizado el acceso a la IA avanzada, permitiendo a investigadores y desarrolladores aprovechar modelos potentes sin necesidad de entrenarlos desde cero.

Los modelos preentrenados han revolucionado campos como el procesamiento del lenguaje natural, llevando a avances en traducción automática, generación de texto, resumen y análisis de sentimientos. También han impulsado progresos en visión por computador, síntesis de voz y otras áreas.

Modelos como **GPT**, **BERT**, y sus sucesores han demostrado capacidades sorprendentes, acercándose en algunos aspectos a la comprensión y

generación de lenguaje humano. Esto ha abierto nuevas posibilidades en interfaces hombre-máquina, asistentes virtuales, y sistemas de IA generativa, al tiempo que plantea importantes cuestiones éticas y sociales sobre el futuro de la IA.

EL QUINTO COMPONENTE: LA GRAN INVERSIÓN EN IA

Entre 2013 y 2023, el volumen de inversión en IA creció exponencialmente. Según estimaciones de *Stanford's AI Index 2023*, la inversión global en IA superó los 200 millardos de dólares en 2021, lo que representa un salto significativo desde los 1,3 millardos de dólares que se invertían anualmente en 2013. Gran parte de esta financiación provino del capital riesgo, dirigido a nuevas empresas de IA y a tecnologías emergentes.

Cifras destacadas:

- **2013-2015:** Las inversiones de capital riesgo en IA eran todavía modestas, pero comenzaron a acelerarse a medida que startups como **DeepMind** (comprada por Google en 2014 por 500 millones de dólares) y **OpenAI** (fundada en 2015) captaron la atención de los inversores.
- **2016-2020:** El capital riesgo se volcó en proyectos de IA, con un aumento espectacular de 26.7 millardos de dólares en 2019, impulsado por tecnologías disruptivas como los vehículos autónomos y las aplicaciones de reconocimiento facial. Los inversores también apostaron por modelos de lenguaje avanzados como GPT-2 y GPT-3, desarrollado por OpenAI.



- **2021-2023:** Hubo un auge de financiación sin precedentes, con 135 millones de dólares invertidos en IA solo en 2021. Las áreas de IA generativa (textos, imágenes, video) recibieron una parte significativa de la inversión, con gigantes tecnológicos y fondos privados dispuestos a liderar esta transformación.

Principales inversores y actores tecnológicos

El papel de los principales actores tecnológicos y fondos de capital riesgo ha sido crucial para este desarrollo. Las grandes tecnológicas, como Google, Microsoft, Amazon, y Meta, junto con firmas de capital riesgo como Sequoia Capital y Andreessen Horowitz, han impulsado el avance de la IA.

- **Google (Alphabet):** ha estado a la vanguardia de la IA desde la adquisición de DeepMind en 2014. Además, invirtió fuertemente en *Google Brain* y en sus modelos de IA de lenguaje, como *BERT* y *PaLM*. El presupuesto de Alphabet destinado a investigación y desarrollo (I+D) en IA superó los veinte millones anuales en 2020.
- **Microsoft:** se ha convertido en un actor clave, con una inversión masiva en **OpenAI**, la compañía detrás de *ChatGPT* y *GPT-4*. En 2023, Microsoft anunció una inversión adicional de diez millones de dólares en **OpenAI**, buscando integrar IA en sus productos de Azure y Office.
- **Amazon:** *Amazon Web Services (AWS)* también ha invertido fuertemente en la nube y soluciones de IA, proporcionando infraestructura para entrenar grandes modelos. AWS desarrolló servicios de IA y aprendizaje automático para empresas, con un presupuesto en IA superior a cuatro millones de dólares en 2022.
- **Meta (Facebook):** ha invertido en IA para mejorar sus sistemas de recomendación, análisis de contenido y experiencias inmersivas en el metaverso. La inversión en laboratorios como *FAIR (Facebook AI Research)* y en el desarrollo de grandes modelos de lenguaje y visión ha sido clave en su estrategia tecnológica.
- **Capital riesgo:** Las firmas de capital riesgo como *Sequoia Capital*, *Andreessen Horowitz*, *Tiger Global*, y *SoftBank* han invertido muchos millones de

dólares en startups de IA. Se calcula que estas firmas gestionaron entre cincuenta y cien millardos de dólares en fondos orientados a la IA solo entre 2020 y 2023. *SoftBank*, por ejemplo, a través de su *Vision Fund*, ha sido responsable de grandes rondas de financiación en startups de robótica e inteligencia artificial.

Áreas clave de inversión en IA

La financiación se ha concentrado en varias áreas clave que han impulsado avances importantes:

- **IA generativa:** *OpenAI*, *Stability AI* y empresas emergentes centradas en modelos de lenguaje y creación de contenidos automatizados.
- **Conducción autónoma:** *Tesla*, *Waymo* (Google) y *Cruise* han recibido cientos de millones para perfeccionar los vehículos autónomos.
- **Robótica e IA aplicada a la manufactura:** Empresas como *Boston Dynamics* y *Nvidia* han obtenido importantes inversiones para desarrollar robots autónomos y hardware especializado para entrenar modelos de IA.
- **Salud:** La IA en salud ha sido un área de alto crecimiento, con empresas como *DeepMind Health* y *PathAI* recibiendo fondos para investigar diagnósticos automatizados y nuevas terapias.

Impacto de la financiación en los avances de la IA

La combinación de capital riesgo y las inversiones de las grandes tecnológicas ha acelerado la evolución de la IA en:

- **Modelos más grandes y potentes:** Como GPT-3 (con 175 millardos de parámetros) y GPT-4, que han abierto nuevas fronteras en procesamiento de lenguaje natural y generación de contenido.
- **IA en el día a día:** Las herramientas de IA se han vuelto omnipresentes en aplicaciones como chatbots, asistentes virtuales (Alexa, Google Assistant), y servicios en línea personalizados.
- **Automatización sectorial generalizada:** Desde el comercio hasta la manufactura y la salud, la IA está automatizando procesos, lo que atrae más inversión y optimiza la productividad.

LA IA GENERATIVA BAJO LA LUPA: ¿HACIA UN NUEVO INVIERNO?

Como se ha visto antes, la evolución de la IA en la última década ha estado marcada por una sinergia entre las enormes inversiones de capital riesgo y las grandes tecnológicas. Este flujo de fondos ha permitido avances rápidos en modelos de IA más potentes, aplicaciones prácticas y nuevos paradigmas tecnológicos, estableciendo las bases para un futuro dominado por la inteligencia artificial.

La carrera por dominar la inteligencia generativa¹⁰ ha resultado en una mejora constante de estos modelos, que siguen expandiendo los límites de lo que es posible en la generación automática de contenido, tanto en texto como en imágenes. Sin embargo, no se puede ignorar la posibilidad de una **burbuja especulativa** en torno a estas tecnologías, impulsada por un **hype recurrente**¹¹. De hecho, analistas como los de **Goldman Sachs**¹² han advertido que la inteligencia artificial podría estar *“sobredimensionada, ser increíblemente costosa y poco confiable”*. Como ha sucedido con otras innovaciones tecnológicas, el entusiasmo desmedido y las promesas a menudo sobrepasan la realidad, lo que plantea la pregunta de si estamos presenciando una revolución duradera o simplemente otra moda pasajera. 📄

10 Entrala, G. [Inspirinas]. (2024, Febrero 12). *El Juego de Tronos de la Inteligencia Artificial* [Video]. YouTube. https://www.youtube.com/watch?v=dULMDS_VJE8

11 Pérez, C. (2004). *Revoluciones tecnológicas y capital financiero: La dinámica de las grandes burbujas financieras y las épocas de bonanza*. México: Siglo XXI Editores.

12 Koebler, J. (2024, julio 12). *Goldman Sachs: AI is overhyped, wildly expensive, and unreliable*. <https://www.404media.co/goldman-sachs-ai-is-overhyped-wildly-expensive-and-unreliable/>

TIPOS DE INTELIGENCIA ARTIFICIAL

💬 César Morillas, Ignacio Despujol y Enric Montesa

TRES CATEGORÍAS DE INTELIGENCIA ARTIFICIAL

Al igual que en otros campos, en el ámbito de la inteligencia artificial (IA) también se ha recurrido al uso de taxonomías y clasificaciones. En algunos casos, se ha intentado organizar los diferentes tipos de inteligencia en función de sus capacidades, de sus aplicaciones o por sus características distintivas. Cada una de estas taxonomías ofrece un enfoque particular para entender el funcionamiento, los límites y el potencial de los sistemas inteligentes.

Como se ha visto en el capítulo anterior, una forma de clasificación de la IA es la utilizada para distinguir la IA de la primera etapa (IA simbólica) de la posterior (IA conexionista), a la que debemos agregar la IA generativa surgida en los últimos años. Pero bajo estas etiquetas, lo que buscan estas clasificaciones no es otra cosa que arrojar luz sobre un campo que se encuentra en rápido desarrollo y expansión.

Entre las diversas clasificaciones que el tiempo ha ido popularizando, la más conocida es la que clasifica a la IA en tres categorías principales:

1. **ANI (Artificial Narrow Intelligence):** o también llamada “IA débil”. Este tipo de IA está diseñada para realizar una tarea muy específica o un conjunto limitado de tareas. Un ejemplo típico es un algoritmo de reconocimiento de imágenes entrenado para identificar gatos en fotografías. Aunque puede ser extremadamente eficiente en esta tarea, su versatilidad es casi nula; no puede realizar tareas fuera de su ámbito estrecho de competencia.
2. **AGI: Inteligencia Artificial General** o “IA fuerte”. Este es el objetivo final de muchos investigadores en el campo de la IA. Se trata de una IA que no solo puede realizar tareas específicas, sino que tiene una versatilidad comparable a la del ser humano. La AGI será capaz de aprender cualquier cosa que un humano pueda aprender, transferir conocimientos de un dominio a otro, y presumiblemente actuar con un alto grado de autonomía.

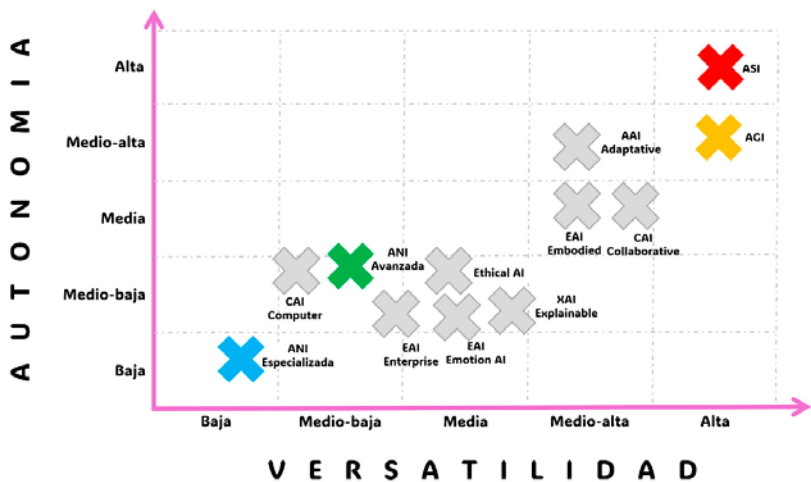
3. **ASI: Inteligencia Artificial Superinteligente.** Este concepto, más especulativo que el anterior, se refiere a la IA que superará en inteligencia a los seres humanos en todos los aspectos, incluyendo la creatividad, la resolución de problemas, la toma de decisiones y la adaptabilidad emocional.

Sin embargo, aunque la clasificación anterior es útil para comprender los diferentes niveles de inteligencia artificial, presenta algunas limitaciones. En particular, no capta con precisión las diferencias intermedias entre distintos tipos de sistemas de IA que, aunque no alcanzan el nivel de una **AGI**, son más versátiles y autónomos que una **ANI**. Tampoco aborda cómo estas capacidades se distribuyen a lo largo de diferentes aplicaciones y dominios de uso.

MAPA DE LAS IA

Es aquí donde el *modelo de versatilidad-autonomía* ofrece una alternativa más matizada y detallada, que permite una comprensión más clara de la diversidad de IA disponibles y sus posibles aplicaciones. Al considerar estos dos ejes como puntos clave, se pueden situar de manera más precisa los diferentes tipos de IA en un continuo, lo que también permite anticipar cómo podrían evolucionar a medida que avancen las investigaciones y desarrollos tecnológicos.

El gráfico siguiente muestra la existencia de un arco de posibilidades desde la Inteligencia Artificial Débil (ANI) hasta la Súper Inteligencia Artificial (ASI).



En el primer caso, la versatilidad y la autonomía de la **ANI** especializada son bajas, mientras que en el segundo, la versatilidad y autonomía de la **ASI** es extremadamente alta (según algunos futurólogos).

VERSATILIDAD

La versatilidad de una inteligencia artificial se refiere a su *capacidad para realizar una amplia variedad de tareas en diferentes dominios o contextos*. Los elementos que definen la versatilidad son aquellos aspectos que permiten a una IA adaptarse a diferentes escenarios, realizar múltiples funciones o resolver problemas en distintas áreas de conocimiento. Según los expertos, la versatilidad de una IA queda definida básicamente por estos elementos:

- ***Dominio de aplicación:*** Se refiere a la capacidad de la IA para operar en un único dominio (estrecho) o en múltiples dominios (amplio).
- ***Capacidad de adaptación:*** Es la habilidad de la IA para ajustar su comportamiento y aprender nuevas habilidades sin requerir reprogramación significativa.
- ***Capacidad multitarea:*** Define si la IA puede manejar varias tareas simultáneamente o si está limitada a una sola tarea a la vez.
- ***Aprendizaje transferido:*** Evalúa si la IA puede aplicar conocimientos o habilidades adquiridos en un contexto a otro diferente.
- ***Contexto en el que opera:*** Se refiere a la capacidad de la IA para funcionar en diferentes entornos o situaciones, desde entornos controlados hasta entornos abiertos y no estructurados.
- ***Profundidad y variedad de habilidades:*** Se refiere a la gama de habilidades que una IA puede ejecutar, desde tareas simples hasta complejas.
- ***Interacción con otros sistemas:*** La capacidad de la IA para integrarse y colaborar con otros sistemas de IA o con humanos, lo que puede aumentar su versatilidad en tareas colaborativas.

En la tabla siguiente se esquematiza lo anterior proporcionando algunos ejemplos que ayudan a ubicar algunas de diferentes IA que aparecen en el gráfico.

Tabla: Elementos que definen la versatilidad en IA con ejemplos

ELEMENTO	DESCRIPCIÓN	EJEMPLO
Dominio de aplicación	Capacidad para operar en uno o varios dominios	ANI Especializada: IA que solo juega al ajedrez
Capacidad de adaptación	Habilidad para aprender y ajustarse a nuevos escenarios o tareas	AAI: Vehículos autónomos que mejoran su conducción al aprender de errores
Capacidad multitarea	Posibilidad de manejar varias tareas simultáneamente	AGI: IA capaz de escribir, programar y resolver problemas matemáticos
Aprendizaje transferido	Aplicar conocimiento de un dominio a otro diferente	Transfer Learning: IA que aprende a traducir textos y luego aplica ese conocimiento a generar subtítulos automáticos
Contexto en el que opera	Capacidad para funcionar en distintos entornos o situaciones	EAI: Robots que operan en fábricas y entornos domésticos
Profundidad y variedad de habilidades	Gama de habilidades que puede ejecutar	ASI: IA capaz de resolver desde problemas de física hasta tareas creativas como pintar
Interacción con otros sistemas	Capacidad de integrarse con otros sistemas o trabajar colaborativamente	CAI: Asistentes médicos que colaboran con doctores para diagnósticos

AUTONOMÍA

La autonomía de una inteligencia artificial se refiere a su *capacidad para operar de manera independiente, tomando decisiones sin intervención humana*. Los elementos que definen la autonomía son aquellos aspectos que permiten a la IA actuar por sí misma en diversas situaciones, evaluar sus entornos, tomar decisiones y ejecutar acciones de manera efectiva y segura.

Elementos que definen la autonomía:

- **Capacidad de toma de decisiones:** La habilidad de la IA para tomar decisiones sin intervención humana, basándose en datos, algoritmos y aprendizajes previos.
- **Independencia operativa:** Se refiere a la capacidad de la IA para realizar tareas sin depender de instrucciones continuas de un humano u otro sistema.
- **Capacidad de autoaprendizaje:** Define si la IA puede mejorar su desempeño aprendiendo de su propia experiencia y retroalimentación sin necesidad de intervención externa.
- **Capacidad de autoevaluación:** La IA debe ser capaz de evaluar su propio rendimiento y corregir errores o ajustar estrategias de forma autónoma.
- **Manejo de incertidumbre:** Es la habilidad de la IA para enfrentar situaciones no previstas o complejas y tomar decisiones bajo incertidumbre.
- **Interacción con el entorno:** Se refiere a la capacidad de la IA para percibir, interpretar y reaccionar a cambios en su entorno de manera autónoma.
- **Nivel de supervisión humana:** Mientras más reducida sea la necesidad de intervención o supervisión humana, mayor será la autonomía de la IA.

La tabla siguiente resume lo expuesto y proporciona algunos ejemplos que ayudan a ubicar algunas de diferentes IA que aparecen en el gráfico.

Tabla: Elementos que definen la autonomía en IA con ejemplos

ELEMENTO	DESCRIPCIÓN	EJEMPLO
Capacidad de toma de decisiones	Habilidad de la IA para decidir sin intervención humana	AGI: Una IA que diagnostica enfermedades sin la ayuda de un médico
Independencia operativa	Capacidad para realizar tareas sin instrucciones continuas	Vehículos Autónomos: Que navegan sin intervención humana en carreteras

Capacidad de autoaprendizaje	Habilidad para aprender de su propia experiencia sin intervención externa	Deep Learning: Algoritmos que ajustan su comportamiento al recibir nueva información
Capacidad de autoevaluación	Evaluar su propio rendimiento y corregir errores o ajustar estrategias	Robots Industriales: Que detectan fallos y ajustan su funcionamiento para mejorar la producción
Manejo de incertidumbre	Capacidad para actuar en situaciones no previstas o complejas	AGI: Resolviendo problemas en tiempo real sin programación específica
Interacción con el entorno	Percibir, interpretar y reaccionar a cambios en su entorno	Robots de Rescate: Que navegan por terrenos complejos y adaptan su comportamiento
Nivel de supervisión humana	Grado de dependencia de intervención o supervisión humana	ASI: Capaz de operar completamente sin supervisión en cualquier entorno

LAS OTRAS IA

En el espacio central del gráfico vemos otras ocho aspas que corresponden a IA con unos niveles de versatilidad y autonomía intermedios. Son las siguientes:

- **CAI** (*Computer-Assisted Instruction* - Instrucción Asistida por Computadora): Se refiere al uso de computadoras para apoyar el proceso educativo. Esto puede incluir software educativo, simulaciones interactivas y plataformas de aprendizaje en línea que ayudan a los estudiantes a adquirir conocimientos y habilidades de manera más efectiva.
- **EAI** (*Enterprise Application Integration* - Integración de Aplicaciones Empresariales): Es el proceso de conectar diferentes aplicaciones empresariales dentro de una organización para que puedan compartir datos y

funciones. Esto ayuda a mejorar la eficiencia y a coordinar las operaciones de manera más fluida.

- **Ethical AI (Inteligencia Artificial Ética):** Se refiere al diseño y desarrollo de sistemas de inteligencia artificial que toman en cuenta principios éticos. Esto incluye garantizar la equidad, la transparencia, la privacidad y la responsabilidad en el uso de IA, así como minimizar los sesgos y evitar daños potenciales a las personas y a la sociedad.
- **XAI (Explainable AI - IA Explicable):** Es un enfoque de la inteligencia artificial que se centra en hacer que los modelos y decisiones de IA sean comprensibles para los humanos. La IA explicativa busca que las decisiones tomadas por los sistemas de IA sean transparentes y que sus procesos sean interpretables, facilitando así la confianza y la capacidad de supervisión por parte de los usuarios.
- **AAI (Autonomous Adaptive Intelligence - Inteligencia Artificial Autónomamente Adaptativa):** Este tipo de IA tiene la capacidad de **aprender y adaptarse** de forma autónoma a través de su experiencia, sin intervención humana significativa. Si bien no llega a los niveles de AGI, la **AAI** se podría situar como un puente entre la ANI y la AGI. Estas IA pueden realizar múltiples tareas y cambiar su comportamiento en función de los resultados de sus interacciones. Los vehículos autónomos y los sistemas de diagnóstico médico adaptativos pertenecen a esta categoría.
- **Inteligencia Artificial Colaborativa (CAI - Collaborative AI)** que aunque no sea necesariamente una categoría distinta en términos de capacidad cognitiva, sino más bien una categoría basada en el modo de interacción con los humanos. Estas IA están diseñadas específicamente para trabajar en conjunto con humanos en tareas que requieren decisiones conjuntas, aprendizaje mutuo y colaboración a largo plazo. Corresponden a este grupo los sistemas de IA en fábricas inteligentes que cooperan con los humanos para realizar tareas en equipo o asistentes médicos que colaboran en el diagnóstico y tratamiento de enfermedades junto con los doctores.
- **Inteligencia Artificial Emocional (Emotion AI o Affective AI)** centrada en la capacidad de las máquinas para reconocer, interpretar y responder a las emociones humanas. Aunque no representa un nivel superior de inteligencia, se puede integrar en diversos tipos de IA, permitiendo interacciones más

- naturales y adaptadas al contexto emocional del usuario. Los asistentes virtuales que detectan emociones en la voz o IA de servicio al cliente que reconocen estados emocionales para adaptar respuestas.
- **Inteligencia Artificial Embodied (EAI - Embodied AI)** referida a la inteligencia artificial que se encuentra integrada en un cuerpo físico (robot o dispositivo) y que interactúa de manera activa con el mundo físico. Estas IA no solo dependen de cálculos o algoritmos, sino que necesitan procesar datos sensoriales y ejecutar acciones en el entorno. Esta capacidad les permite tener una comprensión más profunda y completa de su entorno. Los robots industriales avanzados, los robots de servicio personal y los drones autónomos pueden ser clasificados en este grupo.

En la tabla siguiente aparece el listado con la docena de IA mencionadas y una puntuación (de uno a cinco) en las dimensiones de versatilidad y autonomía.

Listado de IA atendiendo a su gradiente en versatilidad y autonomía (puntos de 1 a 5)

CATEGORÍA	VERSATILIDAD	AUTONOMÍA
<i>ANI: Inteligencia Artificial Estrecha, especializada</i>	1	1
<i>ANI: Inteligencia Artificial Estrecha, avanzada</i>	2	2
<i>AGI: Inteligencia Artificial General o "IA fuerte"</i>	5	4
<i>ASI: Inteligencia Artificial Superinteligente</i>	5	5
CAI Instrucción Asistida por Computadora	2	2
EAI (Integración de Aplicaciones Empresariales)	3	2
Ethical AI: Inteligencia Artificial Ética	3	2
XAI <i>Explainable AI</i> : IA Explicable	3	2
AAI Inteligencia Artificial Autónomamente Adaptativa	4	4
Inteligencia Artificial Colaborativa (CAI - Collaborative AI)	4	3

Inteligencia Artificial Emocional (Emotion AI o Affective AI)	3	3
Inteligencia Artificial Embodied (EAI – Embodied AI)	4	3

INTELIGENCIA HUMANA (IH) Y SINGULARIDAD

La ciencia ficción y los futurólogos se han encargado de difundir la idea de una IA muy avanzada a la que se les ha denominado **Inteligencia General Artificial (AGI)** e **Inteligencia Artificial Superinteligente (ASI)**, dos tipos de IA que en un momento histórico alcanzan a la inteligencia humana o directamente la sobrepasan tanto en versatilidad como en autonomía, entendiendo por esta la ausencia de cortapisas (éticas o morales). Entre los futurólogos y expertos que han escrito sobre este asunto, destacan:

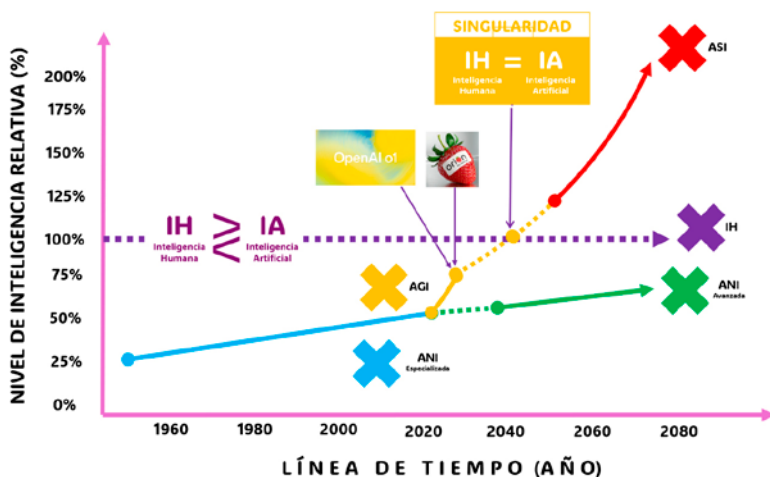
1. **Ray Kurzweil:** sostiene que alcanzaremos la AGI alrededor del año **2029** y asegura que la **ASI** emergerá hacia **2045**, en el contexto de lo que llama la “*singularidad tecnológica*”. En ese punto, la inteligencia artificial superará ampliamente la inteligencia humana, dando lugar a cambios radicales en la sociedad y la tecnología.
2. **Nick Bostrom:** no ofrece una fecha exacta, pero sugiere que la AGI podría ser desarrollada en la **segunda mitad del siglo XXI**, advirtiendo que una vez que se alcance la AGI, la transición a la ASI podría ser rápida, sucediendo **poco después** de la AGI. Para él, la cuestión no es tanto si se alcanzará, sino **cómo se gestionará** el riesgo asociado a la ASI.
3. **Elon Musk:** ha expresado preocupación sobre la rapidez con la que podría desarrollarse la AGI. Aunque no ha dado una fecha precisa, menciona que la inteligencia artificial podría superar a la inteligencia humana entre **2025 y 2030** si no se regula adecuadamente. Ha sido uno de los críticos más vocales sobre los riesgos de la ASI, afirmando que podría representar una amenaza para la supervivencia humana.
4. **Hans Moravec:** pionero en robótica, predijo que **para mediados del siglo XXI** las máquinas alcanzarían un nivel de inteligencia similar al humano. En su libro *Mind Children* (1988), sugiere que el desarrollo de la AGI es inevitable y que, eventualmente, las máquinas reemplazarán a los humanos en

diversas tareas cognitivas. Y aunque no ofrece fechas exactas para la ASI, su visión es que las máquinas superarán a los humanos en todos los aspectos intelectuales, no mucho tiempo después de que alcancemos la AGI.

5. **Vernor Vinge:** matemático y escritor de ciencia ficción conocido por popularizar el concepto de la **singularidad**, en 1993, sugirió que la creación de una inteligencia superior a la humana podría ocurrir **antes de 2030**, con el desarrollo de AGI.
6. **Ben Goertzel:** investigador pionero en AGI y el creador del proyecto *OpenCog*, ha pronosticado que la AGI podría estar entre nosotros hacia **la década de 2020 o 2030**, dependiendo del ritmo de los avances en investigación y desarrollo en IA.
7. **Stuart Russell:** coautor de "Artificial Intelligence: A Modern Approach", tiene una visión más cautelosa. No se compromete con fechas precisas, pero ha advertido que la AGI podría llegar en cualquier momento durante el **siglo XXI**. Su enfoque está en el desarrollo de una **IA ética** y controlable.

Y aunque cada futurólogo tiene una visión única sobre la evolución de la inteligencia artificial, algunos centrándose más en las oportunidades que ofrece, y otros en los riesgos y desafíos que presenta, se puede decir que sus predicciones convergen, siendo las fechas clave para la **AGI** las décadas comprendidas entre **2025 y 2050** y para la **ASI** las décadas comprendidas entre **2040 y 2060**, en función de la velocidad con la que la **AGI** evolucione.

En gráfico siguiente refleja la visión de los autores citados. 📄



03.

FUNDAMENTOS
DE LA
IA

MACRODATOS (BIG DATA)

Interacción entre Datos y Algoritmos
en la Inteligencia Artificial

🗨 Ivana Aguilar

INTRODUCCIÓN

Las nuevas tecnologías introducen términos que, como sociedad, nos vemos en la necesidad de comprender de manera general. Por esto, antes de profundizar en su significado, comencemos con una pregunta básica: ¿A qué llamamos “dato”?

Según Davenport y Prusak (1999) en su libro *“Working Knowledge: How Organizations Manage What They Know”*, un dato es “un conjunto discreto de factores objetivos sobre un hecho real”. Y además de definirlo, los autores lo diferencian del conocimiento y la información.

El término *Big Data* hace referencia a la enorme cantidad de datos digitales, tanto estructurados como no estructurados, que se generan de forma continua. Sin embargo, afirmar que el simple volumen de estos datos aporta valor es incorrecto. El verdadero valor radica en la capacidad de analizarlos e interpretarlos adecuadamente. Para ello, es fundamental utilizar las herramientas adecuadas, que se abordarán en este capítulo. Sin embargo, antes de adentrarnos en las técnicas utilizadas para el procesamiento de datos masivos, es crucial comprender el origen del término Big Data y cómo ha evolucionado hasta convertirse en un elemento clave en el umbral de la emergente Industria 5.0.

Breve historia del Big Data

A finales de 1969 y principios de 1970 surgió la Industria 3.0, conocida como la Revolución Digital. Esta etapa, además de generar una ola de avances tecnológicos, destacó por su principal innovación: la automatización parcial mediante el uso de controles programables y computadoras con capacidad de almacenamiento en memoria. Este cambio transformó radicalmente la producción y mejoró significativamente la eficiencia en las fábricas. Su impacto impulsó la globalización de la economía y dio lugar a nuevas industrias

basadas en las tecnologías de la información y las comunicaciones. Estos avances se iniciaron en los países más industrializados como Estados Unidos y Japón.

Ante el creciente volumen de datos generados por la Industria 3.0, a mediados de la década de 1990, John Mashey, ex científico jefe de Silicon Graphics (SGI), acuñó el término Big Data para describir el manejo y análisis de grandes conjuntos de datos. En su publicación *“Big Data and the Next Wave of InfraStress”*, Mashey demostró una clara comprensión de este fenómeno emergente. A partir de esta idea, SGI adoptó rápidamente el término “Big Data”, utilizándolo como concepto unificador en seminarios y como herramienta publicitaria estratégica.

Evolución de las dimensionalidades en Big Data

En la actualidad, se reconocen hasta 7 dimensiones en Big Data. Este crecimiento en la dimensionalidad refleja el avance continuo de la industria y busca proporcionar un marco más detallado para abordar los desafíos complejos del Big Data, especialmente en lo que respecta a la calidad, consistencia y presentación de los datos.



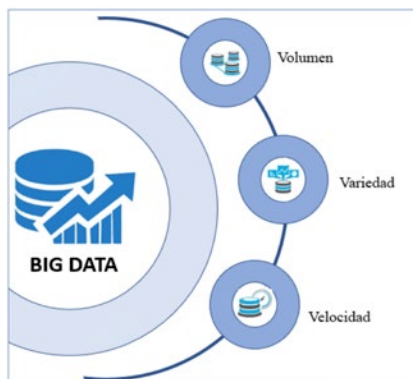
Evolución de las dimensiones en Big Data.

Características iniciales del Big Data: las “3V’s”

A medida que el término Big Data se consolidaba en la industria, se hizo necesario definir las características que debían cumplir los datos para ser considerados como tales.

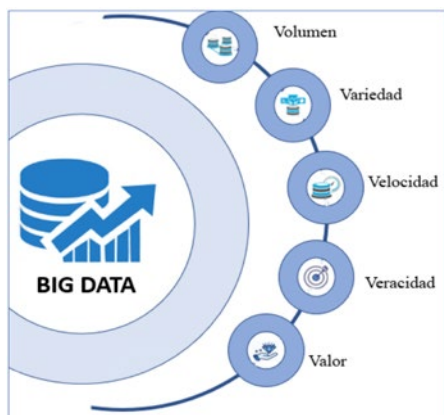
En 2001, Doug Laney, analista de Gartner, publicó el artículo titulado “3D Data Management: Controlling Data Volume, Velocity, and Variety”. En este trabajo, se introdujeron las características iniciales que más tarde se popularizarían como las “3V’s” del Big Data: Volumen, Velocidad y Variedad.

Para ilustrar a qué se refieren estas características tomaremos como ejemplo los datos que maneja una plataforma de *streaming* de video, como Netflix.



- **Volumen:** Netflix maneja enormes volúmenes de datos provenientes de sus usuarios, entre ellos incluye millones de registros de: visualización, búsquedas, calificaciones y recomendaciones. Cada usuario interactúa continuamente con la plataforma, generando una gran cantidad de datos.
- **Velocidad:** Estos datos se generan y se actualizan a alta velocidad. Cada vez que el usuario interactúa con la plataforma ya sea realizando una búsqueda o al término de una serie o película, estos datos deben ser procesados en tiempo real para poder ofrecer recomendaciones instantáneas y actualizaciones en la interfaz de usuario.
- **Variedad:** Los datos en Netflix provienen de diversas fuentes y formatos. Esto incluye datos estructurados, como registros de visualización y calificaciones; y datos no estructurados, como comentarios en redes sociales o reseñas de contenido. Además, la plataforma también debe integrar datos de diferentes dispositivos que puedan utilizar los usuarios, como smartphones, tabletas, ordenadores y televisores inteligentes.

Se puede concluir, con este ejemplo, que la gestión efectiva de las 3V's puede mejorar de forma considerable la experiencia del usuario y, a su vez, ayuda a optimizar los servicios ofrecidos por esta plataforma.



De "3V's" a "5V's"

Esta ampliación de la dimensionalidad en el marco de las "3V's" la estableció nuevamente Douglas Laney, en su artículo *"The Importance of Big Data: A Definition"* publicado en *InformationWeek*, en 2012; en donde, además de las 3 características antes descritas, se añadieron 2 más: la veracidad y el valor.

Laney añadió la veracidad y el valor para abordar cuestiones relacionadas con la calidad de los datos y su capacidad para generar *"insights"* útiles. Esto proporcionaría un enfoque más completo para su gestión y análisis.

Para contextualizar estas dos nuevas características, las incorporaremos a nuestro ejemplo inicial de Netflix.

- **Veracidad:** Es fundamental que Netflix asegure la calidad y precisión de los datos que recopila, ya que los datos inexactos o incompletos pueden llevar a recomendaciones erróneas o a una mala interpretación del comportamiento de los usuarios. Por ejemplo, si un usuario comparte una cuenta con varios miembros de la familia, es importante que los datos sean lo suficientemente fiables para entender correctamente los patrones de visualización y no hacer recomendaciones imprecisas. Por esto, la plataforma, para evitar este tipo de imprecisiones, pone a disposición de los usuarios la posibilidad de elaborar perfiles independientes.
- **Valor:** El valor de los datos radica en la capacidad que tenga la plataforma para transformarlos en *"insights"* útiles. A través de análisis avanzados, Netflix utiliza estos datos para mejorar sus algoritmos de recomendación, personalizar el contenido y mejorar la experiencia del usuario. Esto no solo incrementa la satisfacción de los usuarios, sino que también fortalece las decisiones comerciales, como la producción de contenido original basado en preferencias y tendencias globales.

Con la incorporación de estas dos nuevas dimensiones, resulta evidente que todos los modelos de negocio que implementen Big Data, pueden mejorar sus servicios, garantizando la calidad de los datos y transformándolos en valor tangible tanto para sus usuarios como para la propia empresa.

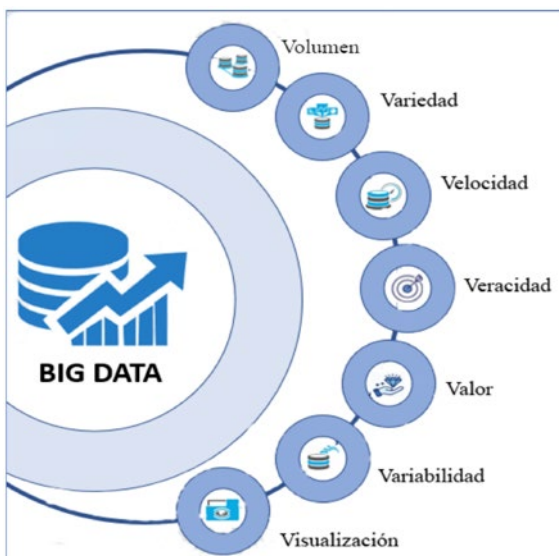
De "5V's" a "7V's"

La idea de las "7V's" comenzó a tomar forma entre 2012 y 2014, sin un autor específico atribuido. Diversas fuentes señalan que fueron expertos en Big Data quienes ampliaron el marco de las "5V's" a medida que el campo evolucionaba, incorporando nuevas dimensiones para abordar aspectos adicionales de la gestión y análisis de datos.

Además de las "5V's", la variabilidad y la visualización, fueron las dimensiones que se integraron en diversos estudios para proporcionar un enfoque más completo en el manejo y análisis de datos complejos. Estas dos nuevas dimensiones permiten abordar la inconsistencia en los datos y mejorar su interpretación a través de representaciones visuales efectivas.

Añadiremos estas 2 nuevas dimensionalidades con el propósito de ilustrar y finalizar nuestro ejemplo inicial: Netflix.

- **Variabilidad:** Los patrones de visualización de los usuarios pueden variar considerablemente a lo largo del tiempo. Por ejemplo, el comportamiento de los usuarios puede cambiar según la temporada, las tendencias o eventos especiales. Es por esto que la plataforma debe ser capaz de adaptarse a



estos cambios en los datos para mantener la precisión de sus recomendaciones y análisis.

- **Visualización:** Es sabido que Netflix utiliza herramientas avanzadas de visualización de datos para interpretar grandes volúmenes de información y presentarla de manera clara a los equipos internos. Esto le permite identificar patrones de consumo, tendencias globales y puntos clave para tomar decisiones estratégicas. Estas visualizaciones también facilitan el monitoreo de métricas como la retención de usuarios o la popularidad de determinados títulos.

Como conclusión a este bloque podemos afirmar que cada dimensión juega un papel clave en la gestión y análisis de grandes conjuntos de datos, permitiendo a los modelos de negocio no solo manejar grandes cantidades de información, sino también garantizar su calidad, adaptarse a cambios, extraer valores estratégicos y presentar los resultados de manera comprensible. Con la adopción de este marco, las empresas pueden tomar decisiones más informadas, optimizar procesos y mejorar tanto la experiencia del usuario como su rendimiento comercial.

FUENTES DE DATOS EN BIG DATA

Existen numerosas fuentes de datos que hacen posible el funcionamiento de toda la maquinaria del *Big Data*, las cuales se pueden clasificar según diversas características como su origen o tipo de acceso, es decir, si son de acceso gratuito o de pago.

Entre las más importantes, destacan las siguientes:

- **Las redes sociales (RRSS):** Facebook (ahora Meta) como red social pionera fue una de las primeras en explotar esta idea de negocio. Mark Zuckerberg, CEO de Meta, en 2018 durante una audiencia en el Congreso de los Estados Unidos, al ser preguntado sobre los datos recopilados por su plataforma y sobre qué iba a hacer con ellos, explicó que era un servicio gratuito, en donde los usuarios compartían información de forma voluntaria. Entre la información compartida por los usuarios de la plataforma se encuentran: publicaciones, fotos y mensajes, además de

ciertos datos técnicos como direcciones IP, información sobre los dispositivos utilizados y la ubicación aproximada del usuario. Zuckerberg, para concluir, mencionó que Facebook utiliza esta información para mejorar la experiencia del usuario, ofrecer anuncios personalizados y asegurar la seguridad de la plataforma.

- Otras redes sociales, Instagram, TikTok, X (antes Twitter), han adoptado este mismo modelo de negocio, basado en la recopilación y utilización de datos.
- **Sensores IoT (Internet de las cosas):** Recopilan información en tiempo real desde diversos entornos, por ejemplo: ciudades, fábricas, vehículos y hogares, generando de esta forma una gran cantidad de datos. La capacidad de estos dispositivos para recolectar datos de manera continua y automatizada permite obtener información precisa y actualizada, lo que facilita su análisis y agiliza la toma de decisiones en áreas como la optimización de procesos, la predicción de fallos y el desarrollo de servicios personalizados. Entre los datos que estos sensores capturan se incluyen métricas como temperatura, humedad, movimiento, ubicación geográfica y niveles de uso, entre otros.
- **Bases de datos:** Las bases de datos relacionales (que utilizan SQL) y las bases de datos no relacionales (NoSQL) están consideradas entre las principales fuentes de datos debido a su impresionante capacidad de almacenamiento. Mientras que las bases de datos relacionales son ideales para gestionar datos estructurados y relaciones complejas, las bases de datos no relacionales ofrecen flexibilidad y escalabilidad para manejar grandes volúmenes de datos no estructurados.
- **Transacciones comerciales:** Los bancos, en su actividad diaria, generan grandes volúmenes de datos relacionados con las transacciones y comportamientos financieros de sus clientes. Esta información refleja las actividades económicas de los usuarios, como movimientos bancarios, inversiones, compras y pagos, proporcionando una valiosa fuente de datos para análisis y toma de decisiones estratégicas.
- **Sistemas de información empresarial:** Tanto los CRM (sistemas de gestión de relaciones con clientes) como los ERP (sistemas de gestión empresarial) ofrecen información valiosa. Los CRM, en particular, facilitan la mejora de la relación con los clientes al proporcionar datos detallados sobre interacciones,

preferencias y comportamiento, mientras que los ERP optimizan la gestión interna de la empresa al integrar y analizar datos de diferentes áreas operativas.

TÉCNICAS Y HERRAMIENTAS DEL BIG DATA

Antes de abordar las técnicas y herramientas utilizadas en Big Data, es fundamental entender la diferencia entre estos dos conceptos que, aunque distintos, son complementarios entre sí.

Las **técnicas**, hacen referencia a los métodos y enfoques utilizados para analizar y procesar grandes volúmenes de datos. Aquí se incluyen: algoritmos, modelos y procedimientos que permiten extraer información útil y tomar decisiones basadas en datos.



Por otro lado, las **herramientas**, son aplicaciones, plataformas y software que facilitan la implementación de las técnicas utilizadas en Big Data. Estas herramientas permiten gestionar, procesar, analizar y visualizar datos de manera eficiente.

A continuación, se muestran las etapas de tratamiento por las que pasan estas grandes cantidades de datos.

En la primera etapa de tratamiento del Big Data, tenemos el **almacenamiento**, que es el contenedor o infraestructura que nos va a permitir gestionar y procesar estos datos. Entre algunas de las principales opciones de almacenamiento tenemos: el almacenamiento en la nube, sistemas de almacenamiento distribuido, SQL y NoSQL, **data warehouses** y sistemas de almacenamiento en disco local o NAS.

A continuación, tenemos la etapa del **procesamiento** de datos. Es en esta etapa en donde entran en juego las distintas técnicas que en la actualidad se usan en Big Data. Algunas de las técnicas de uso común son: la limpieza de datos, la transformación de datos, minería de datos, machine learning (ML), procesamiento del lenguaje natural (NPL), análisis predictivo, análisis de series temporales, análisis de clústeres, análisis de redes, entre otras.

Por último, está la etapa del **análisis** de datos, donde las técnicas se integran con herramientas especializadas para implementar algoritmos y realizar análisis avanzados. En esta fase, se emplean plataformas desarrolladas por grandes empresas tecnológicas como IBM, Microsoft, Oracle y Google, que proporcionan soluciones potentes para manejar la complejidad del análisis de grandes volúmenes de datos.

Entre las herramientas más destacadas tenemos: **IBM** (*IBM Watson Studio* e *IBM InfoSphere BigInsights*), **Microsoft** (*Azure HDInsight* y *Azure Synapse Analytics*), **Oracle** (*Oracle Big Data Service* y *Oracle autonomous Data Warehouse*), **Google** (*BigQuery* y *Google Cloud Dataflow*).

APLICACIONES Y USOS DEL BIG DATA

El Big Data ofrece innumerables aplicaciones que impactan en diversas industrias, desde la optimización de procesos hasta la personalización de servicios. A continuación, se presentan algunos de los principales usos de esta tecnología en diferentes campos.

- **Negocios y marketing:** el Big Data se utiliza para analizar grandes volúmenes de datos de clientes, identificar patrones de comportamiento y prever tendencias del mercado. Esto permite a las empresas personalizar campañas publicitarias, optimizar estrategias de ventas y tomar decisiones basadas en datos precisos para maximizar el rendimiento y la satisfacción del cliente.
- **Salud:** El análisis de datos colectivos puede ser clave en el desarrollo de diversas herramientas, destacando entre ellas los sistemas AID (Aided Diagnosis) o sistemas de apoyo al diagnóstico. Estos sistemas permiten a los profesionales de la salud ingresar características clínicas para obtener diagnósticos diferenciales precisos y relevantes. Al hacerlo, se mejora la

atención al paciente, ya que le ofrece una mayor personalización en el cuidado de su salud.

- **Sector público:** A partir de datos recopilados mediante sensores *IoT* o por cámaras de tráfico, es posible desarrollar aplicaciones que ofrezcan consultas en tiempo real para optimizar el tráfico, predecir los tiempos de llegada a destinos específicos y proporcionar información precisa y relevante para los turistas.
- **Ciencia y tecnología:** La recopilación de datos en estos campos impulsa nuevas líneas de investigación en ciencia y tecnología, facilitando el desarrollo de dispositivos más avanzados y alineados con los avances científicos. Esto permite una continua actualización y mejora de las soluciones tecnológicas.

DESAFÍOS DEL BIG DATA

A pesar de los numerosos beneficios que ofrece el Big Data, todavía existen importantes desafíos que deben superarse. A continuación, se destacan los tres más relevantes, los cuales, a medio plazo continuarán generando más de un quebradero de cabeza.

- **Privacidad y seguridad:** ¿Qué tan seguras son las plataformas, dispositivos o aplicaciones en las que compartimos nuestros datos? La realidad es que la **seguridad absoluta** en informática no existe. Ante la necesidad urgente de regular tanto el flujo como el tratamiento de los datos recopilados de diversas fuentes, la Unión Europea (UE) implementó en 2018 el Reglamento General de Protección de Datos (GDPR, por sus siglas en inglés). Este reglamento establece estrictas normativas sobre cómo las empresas y organizaciones deben gestionar los datos, con el objetivo principal de proteger la privacidad de los ciudadanos de la UE. Entre los aspectos más relevantes del GDPR en relación con el Big Data se incluyen: el consentimiento explícito de los usuarios, el derecho al olvido, la transparencia y acceso a los datos, la seguridad en su manejo, y la obligación de notificar cualquier violación de seguridad. Este marco regulatorio ha tenido un impacto significativo en la gestión de Big Data en Europa,

obligando a las empresas a tratar la información de manera responsable y garantizando la protección de los ciudadanos.

- Además, la UE continúa desarrollando nuevas políticas y directrices para adaptarse a los avances tecnológicos, especialmente con la llegada de la Industria 5.0.
- **Calidad de los datos:** Garantizar la calidad de los datos en Big Data es fundamental. Para ello, es necesario implementar sistemas y procesos automatizados que permitan la limpieza, validación y verificación de los datos de manera eficiente. Esto implica utilizar técnicas como la deduplicación, imputación de valores faltantes, integración de datos y monitoreo continuo. Además, podemos decir que las herramientas de análisis deben estar capacitadas no solo para gestionar grandes volúmenes de información, sino también para abordar las complejidades asociadas a la diversidad de formatos, la velocidad de procesamiento y la precisión de los datos.
- **Costes de infraestructuras:** Representan uno de los factores más críticos en las empresas ya que estos costes no solo incluyen el almacenamiento y procesamiento de datos, sino también el mantenimiento de sistemas y tecnologías avanzadas necesarias para gestionar la información de manera eficiente y segura. Algunos de los componentes que podemos identificar entre estos costes son: el almacenamiento de datos, el procesamiento de datos, infraestructura en la nube vs. On-premise, costes de seguridad y cumplimiento, personal y talento especializado y, por último, el mantenimiento y actualizaciones de estos.

FUTURO DEL BIG DATA

En la introducción de este capítulo hemos mencionado el impacto que tendrá el Big Data en la emergente Industria 5.0, donde la colaboración entre humanos y máquinas será aún más estrecha. Acompañado de la inteligencia artificial, el Big Data impulsará nuevas tendencias innovadoras que transformarán profundamente la sociedad. A continuación, exploraremos brevemente cómo estas tecnologías moldearán el futuro y sus principales contribuciones en estos sectores.

- **Inteligencia Artificial (IA) y Big Data:** El Big Data será fundamental para el avance de la inteligencia artificial, ya que, al suministrar grandes volúmenes de datos a los modelos de IA, estos podrán mejorar en su entrenamiento y precisión. Este enriquecimiento de datos facilitará el desarrollo de aplicaciones más sofisticadas en áreas como el procesamiento de lenguaje natural y la visión por computadora, proporcionando la información necesaria para identificar patrones y tomar decisiones más precisas. En resumen, la integración de Big Data e IA potenciará la creación de soluciones más inteligentes y efectivas.
- **Big Data en la sociedad:** Aunque la seguridad y privacidad de los datos siguen siendo uno de los mayores desafíos en el ámbito del Big Data, abordarlos es necesario para maximizar sus beneficios de forma ética y equitativa. A pesar de estos retos, el Big Data ya está impactando positivamente en la sociedad, impulsando avances significativos en diversos sectores como: la salud, la educación y el transporte, entre otros. Su capacidad de procesar y analizar grandes volúmenes de información permite una toma de decisiones más informada en negocios y política, además de contribuir a una mejor gestión de los recursos naturales y medioambientales.
- **Tendencias emergentes:** Muchas son las tendencias que van abriéndose paso conforme evoluciona en la actualidad el Big Data. Algunos de estos principales avances se centran en: la analítica predictiva y prescriptiva, el edge computing, la democratización de los datos, el Blockchain para la integridad de los datos, la integración multicanal y la analítica en tiempo real. 📄

REDES NEURONALES ARTIFICIALES

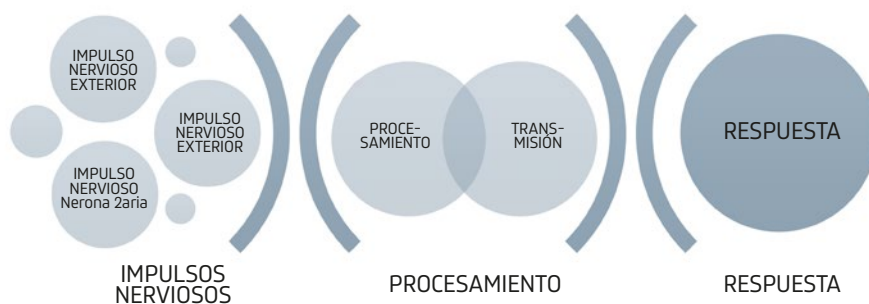
🗨️ Jesús Alfaro y Enric Montesa

REDES NEURONALES BIOLÓGICAS Y ARTIFICIALES:

¿ANALOGÍA O INGENIERÍA INVERSA?

Una **red neuronal artificial (RNA)** es un modelo de aprendizaje compuesto por capas y nodos que simula el funcionamiento del cerebro humano. A través del entrenamiento, adquiere memoria y mejora continuamente su capacidad para resolver problemas, de menor a mayor complejidad¹.

Las redes neuronales artificiales se inspiran en las biológicas, que reciben impulsos nerviosos, los procesan y generan una respuesta. Los neurocientíficos descomponen y estudian el cerebro, que tiene miles de millones de neuronas interconectadas, para entender cómo estas redes biológicas se comunican, procesan y almacenan información. Este conocimiento ha servido de base para diseñar redes neuronales artificiales.

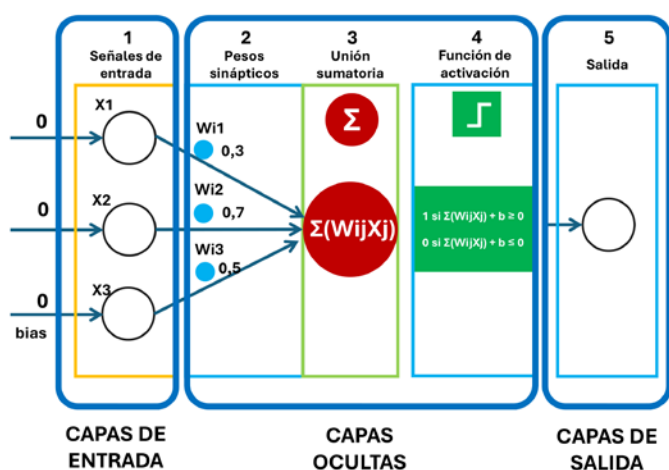


Esquema conceptual de una red neuronal biológica

- 1 Franco Guzmán, M., Romero Romo, M. A., Palomar Pardavé, M. E., Cobos Murcia, J. Á., Álvarez Romero, G. A., & Hernández Ramírez, D. *De neuronas biológicas a neuronas artificiales: El fascinante mundo de las redes neuronales*. Universidad Autónoma del Estado de Hidalgo. Disponible en <https://www.uaeh.edu.mx/divulgacion-ciencia/redes-neuronales/>

Esquema conceptual de una red neuronal artificial

Las redes neuronales artificiales reciben datos de entrada (imágenes, texto, audio, etc.) y los ponderan con un peso, que varía con el entrenamiento. Utilizan funciones de activación y algoritmos de procesamiento para generar una respuesta de salida. Durante el entrenamiento, el algoritmo de *retropropagación* ajusta el peso y las funciones de activación, minimizando el error entre la salida real y la deseada, lo que genera aprendizaje. Este proceso se asemeja a la plasticidad cerebral.



Esquema conceptual de una red neuronal artificial.

La mejor forma de entender el campo de las redes neuronales artificiales es asumir que están basadas en la **ingeniería inversa** del cerebro humano, que descompone y analiza el funcionamiento de las redes neuronales biológicas, lo cual ha facilitado la replicación de los mecanismos cerebrales en sistemas computacionales aplicados a la IA. El *aprendizaje profundo* también se ha inspirado en la organización jerárquica del cerebro al utilizar múltiples capas interconectadas para procesar información compleja.

En definitiva, la IA conexionista basada en el aprendizaje y las RNA es el resultado de la aplicación de la ingeniería inversa a las redes biológicas.



HISTORIA, EVOLUCIÓN Y TIPOLOGÍA DE REDES NEURONALES ARTIFICIALES

Los Inicios: McCulloch y Pitts (1943)

McCulloch y Pitts propusieron una red simplificada de neuronas artificiales que realizaba cálculos booleanos (verdadero/falso). Aunque muy básica, esta idea abrió la puerta a la noción de que el cerebro puede ser modelado como un sistema computacional y que el procesamiento cognitivo podría emularse en máquinas.

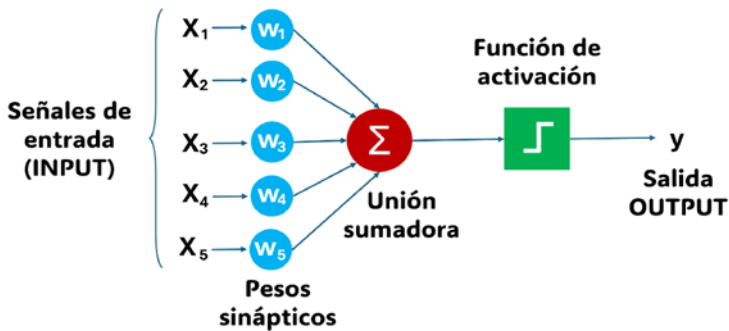
En este modelo, cada *neurona artificial* se comportaba de manera binaria: si recibía suficientes entradas activas (similar a las sinapsis en una neurona biológica), producía una salida; de lo contrario, permanecía inactiva. Esta idea fue extremadamente influyente, ya que ofreció una base para las futuras redes neuronales y sus aplicaciones en la computación.

El Perceptrón: Frank Rosenblatt (1958)

Una década después fue cuando un psicólogo e investigador de la informática desarrolló en 1958 el primer modelo de red neuronal práctico conocido como Perceptrón, un sistema que intentaba aprender automáticamente a partir de datos. Frank Rosenblatt orientó su invento hacia la clasificación de patrones resolviendo el problema de cómo diferenciar entre imágenes de letras o números.

En esa primitiva red neuronal ya estaba casi todo: una capa de *entradas* (características o atributos de un problema), una *capa de pesos* (que determina la importancia de cada entrada), y una salida (la decisión final, que puede ser clasificar algo como 1 o 0, por ejemplo). Además, el sistema ajustaba los pesos de las entradas a través de un proceso de aprendizaje, lo que le permitía mejorar su capacidad de clasificación con el tiempo.

Era un tipo de red neuronal muy simple, y su estructura se limitaba a una sola capa de neuronas, tal y como se puede observar en el gráfico siguiente y se explica en el cuadro sobre el funcionamiento del perceptrón.



FUNCIONAMIENTO DEL PERCEPTRÓN

El Perceptron es una unidad que simula el comportamiento de una neurona biológica, tomando varias entradas, procesándolas y generando una salida. Es especialmente útil en tareas de clasificación binaria, es decir, cuando se quiere determinar si una entrada pertenece a una clase o a otra.

Los componentes básicos del Perceptron son los siguientes:

1. **Entradas:** El perceptrón recibe varias entradas, representadas como $x_1, x_2, \dots, x_n$. Cada entrada puede ser una característica de los datos que estamos analizando. Por ejemplo, si queremos clasificar si un correo es spam o no, las entradas podrían ser palabras, la longitud del correo, etc.
2. **Pesos:** A cada entrada se le asigna un peso $w_1, w_2, \dots, w_n$. Los pesos determinan la importancia de cada entrada. Inicialmente se establecen de manera aleatoria y se ajustan durante el entrenamiento.
3. **Suma ponderada:** El perceptrón calcula una suma ponderada de las entradas multiplicando cada entrada por su peso correspondiente:

$$z = w_1x_1 + w_2x_2 + \dots + w_nx_n + b = w_1x_1 + w_2x_2 + \dots + w_nx_n + b$$
4. Aquí b es el **sesgo** (bias), que actúa como un valor ajustable para desplazar la función de activación y controlar mejor la salida del perceptrón.

- 5 . **Función de activación:** La suma ponderada pasa por una **función de activación**. En el perceptrón simple, esta función es una **función escalón**, que decide si la salida es 0 o 1, indicando si la entrada pertenece a una clase o a otra.
- 6 . **Entrenamiento:** Durante el entrenamiento, el perceptrón ajusta los pesos w para reducir los errores de clasificación. Utiliza el algoritmo de aprendizaje del perceptrón, que actualiza los pesos en función de la diferencia entre la salida esperada y la salida real.

Críticas al Perceptron y temprana desaparición de Rosenblatt

Aunque inicialmente generó entusiasmo por su capacidad de aprendizaje, las críticas no tardaron en surgir, especialmente tras la publicación del libro *Perceptrons* en 1969 por Marvin Minsky y Seymour Papert². En esta obra, los autores demostraron las limitaciones del perceptrón, particularmente su incapacidad para resolver problemas no lineales, como la operación XOR, sin una capa oculta. Esta crítica influyó en la comunidad científica, que se distanció del enfoque de las redes neuronales durante casi dos décadas. Además, la temprana muerte de Rosenblatt en 1971 a los 43 años también contribuyó a que su visión no avanzara como podría haberlo hecho, dejando un legado truncado en el desarrollo inicial de la inteligencia artificial.

El Renacimiento de las Redes Neuronales:

Redes Multicapa (1980-1990)

A pesar del escepticismo, las redes neuronales no desaparecieron. En la década de 1980, los investigadores comenzaron a explorar redes neuronales más complejas, específicamente las *redes multicapa* o redes neuronales de *propagación hacia adelante*. Estas redes contenían múltiples capas de neuronas, lo que permitía resolver problemas no lineales que el perceptrón no podía manejar.

El avance crucial en este periodo fue el algoritmo de *retropropagación* del error (o *backpropagation*), popularizado por Geoffrey Hinton (premio nobel de física 2024 por sus avances en redes neuronales artificiales), David

2 Minsky, M., & Papert, S. (1988). *Perceptrons: An Introduction to Computational Geometry (Expanded ed.)*. Cambridge, MA: MIT Press.

Rumelhart y Ronald Williams en 1986³. La *retropropagación* permite que las redes ajusten los pesos de las neuronas de manera eficiente, minimizando el error de predicción mediante el cálculo de gradientes y optimizando a través de múltiples capas. Esto resolvió el problema clave del perceptrón simple, ya que ahora las redes neuronales podían aprender y ajustar sus parámetros en capas ocultas.

Este nuevo enfoque abrió las puertas a modelos mucho más potentes que podían aprender representaciones más complejas de los datos. Durante los años 80 y 90, las redes neuronales empezaron a utilizarse en áreas como el reconocimiento de patrones, la clasificación de imágenes y la predicción de series temporales.

Redes Neuronales Profundas:

El Surgimiento del “Deep Learning” (2000-2010)

Aunque las redes multicapa ofrecían un avance significativo, el verdadero potencial de las *redes neuronales profundas* (*Deep Learning*) no se realizó hasta las décadas de 2000 y 2010. Como se verá en un capítulo posterior de la Guía, el *Deep Learning* se refiere a redes neuronales con muchas capas ocultas, que permiten la extracción automática de características de datos complejos. Estas redes son capaces de descubrir representaciones jerárquicas de los datos, lo que significa que las capas más bajas aprenden características simples (como bordes en una imagen), mientras que las capas superiores combinan estas características para detectar patrones más complejos (como la forma de un objeto)⁴.

Una de las razones principales por las que el *Deep Learning* ganó popularidad en esta época fue el aumento en la potencia de cálculo, particularmente a través del uso de *Unidades de Procesamiento Gráfico* (GPU) que aceleraron el entrenamiento de estas redes. Al mismo tiempo, grandes cantidades de

3 Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). *Learning representations by back-propagating errors*. *Nature*, 323(6088), 533–536. <https://doi.org/10.1038/323533a0>

4 Berzal, Fernando (2019): *Redes Neuronales & Deep Learning - Volumen 1: Entrenamiento de redes neuronales artificiales [formato 8.5" x 11"]*, Independently published, ISBN 10: 1090320302 / ISBN 13: 9781090320308, p. XX.

datos (big data) se volvieron disponibles, lo que permitió entrenar redes más grandes y complejas.

Los pioneros del Deep Learning, como Geoffrey Hinton, Yann LeCun y Yoshua Bengio⁵, desarrollaron arquitecturas más avanzadas de redes neuronales, como las *Redes Neuronales Convolucionales* (CNN) y las *Redes Neuronales Recurrentes* (RNN).

Redes Neuronales Convolucionales (CNN)

Las CNN fueron revolucionarias en el campo del reconocimiento de imágenes. Inspiradas en la estructura del cerebro, las CNN utilizan *convoluciones* para detectar características locales en imágenes, lo que las hace ideales para tareas como el reconocimiento facial, la segmentación de imágenes y la visión por computadora en general. Yann LeCun fue un pionero en este campo, y su red LeNet sentó las bases para modelos más avanzados.

Redes Neuronales Recurrentes (RNN)

Por otro lado, las RNN fueron diseñadas para manejar datos secuenciales, como el lenguaje o las series temporales. A diferencia de las redes de propagación hacia adelante, las RNN tienen conexiones recurrentes que permiten mantener una “memoria” de las entradas anteriores, lo que es útil en aplicaciones como el procesamiento del lenguaje natural y el reconocimiento de voz. Las LSTM (Long Short-Term Memory) y las GRU (Gated Recurrent Unit) son variantes de RNN que han sido fundamentales para superar problemas de memoria a corto plazo en redes neuronales.

Redes Neuronales Modernas:

Transformadores y Modelos Generativos

El siguiente gran avance en el campo de las redes neuronales fue el desarrollo de la arquitectura *Transformer*, introducida en 2017 en el artículo “*Attention*

5 Fundación Princesa de Asturias. (2022, junio 16). *Geoffrey Hinton, Yann LeCun, Yoshua Bengio y Demis Hassabis, Premio Princesa de Asturias de Investigación Científica y Técnica 2022*. Fundación Princesa de Asturias. <https://www.fpa.es/es/premios-princesa-de-asturias/premiados/2022-geoffrey-hinton-yann-lecun-yoshua-bengio-y-demis-hassabis.html>

is All You Need” por investigadores de Google⁶. Los transformadores fueron diseñados originalmente para mejorar el procesamiento del lenguaje natural (PLN), pero han demostrado ser extremadamente versátiles.

Transformadores y Atención

El *transformer* se basa en un mecanismo de atención, que permite que la red se enfoque en diferentes partes de una secuencia de datos de manera dinámica. Esto ha resultado ser muy efectivo para tareas como la traducción automática, el análisis de texto y la generación de lenguaje natural. Un modelo de transformers aprende a relacionar diferentes partes de una secuencia de manera mucho más eficiente que las RNN, lo que ha llevado a avances en la comprensión y generación de lenguaje.

Modelos como *GPT (Generative Pretrained Transformer)* de OpenAI han demostrado capacidades asombrosas en la generación de texto coherente, ya que pueden predecir y generar secuencias largas de palabras con una fluidez similar a la humana. Estos avances han hecho posible la creación de sistemas de IA que pueden interactuar con humanos a través de lenguaje natural⁷.

Inteligencia Artificial Generativa Multimodal

La IA generativa no se limita solo al lenguaje. Modelos más recientes han dado lugar a *inteligencia artificial multimodal*, que puede procesar y generar múltiples tipos de datos, como texto, imágenes y sonidos. *DALL-E*, por ejemplo, es un sistema que genera imágenes a partir de descripciones textuales, utilizando una combinación de modelos de transformers y redes convolucionales. Otro ejemplo es *CLIP*, que relaciona imágenes y texto de manera semántica.

6 Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008. <https://doi.org/10.48550/arXiv.1706.03762>

7 Ver en la Guía los apartados dedicados al Procesamiento del Lenguaje Natural (PLN) y los Grandes Modelos de Lenguaje (LLM)



“Genera una ilustración apaisada y visualmente atractiva que demuestre el concepto de capacidades de IA multimodal. La imagen muestra un espacio de trabajo futurista con una pantalla ancha de ordenador que despliega elementos interconectados: texto (como un fragmento de documento), imágenes (una pintura vibrante y un diagrama técnico) y formas de onda de audio. Sobre la pantalla,

una proyección holográfica muestra una IA procesando estos elementos simultáneamente. El espacio de trabajo es moderno y elegante, con una iluminación sutil. En la pantalla y en el holograma aparece el texto 'Made by DALL-E' de forma destacada para enfatizar el papel de la IA. El estilo debe ser limpio, profesional y educativo, resaltando el potencial de la IA para incrementar la productividad.”

Con este prompt, se puede comprender cómo solicitar a DALL-E una imagen como la que aparece en la parte superior

Estos modelos multimodales son una evolución de las redes neuronales profundas y han sido posibles gracias a la integración de técnicas avanzadas de *aprendizaje no supervisado* y la enorme cantidad de datos disponibles para entrenar redes neuronales cada vez más grandes.

APLICACIONES PRÁCTICAS DE LAS REDES NEURONALES A LA IA

Existen numerosas aplicaciones prácticas de las redes neuronales a la IA. Pero hemos querido resaltar las siguientes:

- **Visión artificial:** las neuronas tipo CNN han provocado una evolución de la visión artificial, permitiendo la detección de rostros, segmentación de imágenes y reconocimiento de imágenes.
- **Procesamiento del lenguaje natural (Natural Language Processing, PLN):** Las RNN se utilizan en tareas de traducción automática, generación de textos en diferentes idiomas. Su aplicación al reconocimiento de la voz y a los **chatbots** como Siri (Apple), Alexa (Amazon) y Cortana (Microsoft) han revolucionado el lenguaje automático por voz. Estos sistemas son capaces de procesar la señal de voz, transformarla en texto o en órdenes para dirigir una máquina. Otro algoritmo desarrollado en el marco del desarrollo de PLN es el conocido **BERT** (Bidirectional Encoder Representations from Transformers), que es un algoritmo utilizado por Google para la mejora del lenguaje de los usuarios a la hora de realizar búsquedas complejas.
- **Diagnóstico médico por imagen:** el desarrollo de las CNN ha implicado una mejora en el ámbito médico, mejorando la diagnosis de numerosas patologías que requieren de especialistas cualificados con cierta experiencia. Un ejemplo es la detección de tumores cerebrales: *"...Una de estas soluciones consiste en la creación de redes neuronales convolucionales que consigan auto aprender patrones y algoritmos para detectar y clasificar de forma automática tumores cerebrales. Además, el punto diferencial de esta solución es que la red creada, se alimenta con una base de datos basada en imágenes de resonancia magnética (IRM) de pacientes de todo el mundo".*
- **Conducción autónoma:** El Deep learning supuso una revolución en campos como sistemas de conducción autónoma. Estas redes neuronales profundas, junto con las CNN, permitieron la toma de decisiones del tipo, cuando debe parar y cuando debe continuar la marcha un vehículo, mediante respuestas sencillas que constituyen la base de la conducción autónoma actual.
- **Desarrollo de modelos de IA generativa:** un ejemplo del desarrollo de redes neuronales profundas lo tenemos en la mejora de los prompts que permiten optimizar el rendimiento de las tecnologías avanzadas de IA generativa en entornos conversacionales. Ejemplos de este tipo los tenemos en **ChatGPT** y **DALL_E** de Open AI, **GEMINI Pro** de Google (antiguo **BARD**), **CLAUDE 3** de Anthropic, entre otros... 

APRENDIZAJE AUTOMÁTICO (ML: MACHINE LEARNING)

La ciencia detrás de la magia

🗨 Ignacio Despujol Zabala

En este capítulo trataremos el aprendizaje automático (machine learning en inglés), una rama de la inteligencia artificial que ha revolucionado la forma en que abordamos problemas complejos y procesamos grandes cantidades de datos y que está en la base de los grandes avances en inteligencia artificial generativa, tan de moda últimamente. Sus orígenes se remontan a mediados del siglo XX, aunque su auge y aplicación generalizada son mucho más recientes.

El concepto de aprendizaje automático surgió en la década de 1950, cuando los pioneros de la disciplina comenzaron a explorar la idea de que las computadoras pudieran aprender sin ser programadas explícitamente para cada tarea. Hasta ese momento las ciencias de la computación se basaban en desarrollar algoritmos que indicarán a los ordenadores lo que tienen que hacer exactamente en cada momento y para cada situación posible (que es lo que se sigue haciendo en la mayoría de los programas que usamos en nuestro día a día en ordenadores, tablets, teléfonos y dispositivos electrónicos diversos).

Lo primero es entender que un algoritmo en realidad no es más que una serie de pasos definidos para resolver un problema. Como una receta de cocina, vaya. Si sigues los pasos en orden, al final obtienes el resultado que buscas, pero si te saltas alguno o aparece un imprevisto, el resultado puede ser completamente distinto de lo esperado, ya que los pasos a seguir están "escritos en piedra". El cambio de paradigma en el caso del aprendizaje automático es que los algoritmos que se crean no son las instrucciones para resolver el problema a tratar, sino para analizar datos existentes y extraer patrones que permitan crear un modelo que nos dé soluciones al problema, incluso para casos que no se habían contemplado al crearlos.

En este punto es conveniente introducir el concepto de modelo. Un modelo es una representación simplificada de la realidad que los seres humanos creamos para hacer manejable la complejidad del mundo que nos rodea, de forma que nos ayuden a entender y predecir fenómenos complejos. La clave

está en simplificar lo suficiente para poder trabajar con el modelo, pero no tanto como para perder los detalles cruciales que queremos estudiar.

Un modelo simple podría ser: "Los aparatos diseñados por el hombre que vuelan son aviones." Este modelo tiene el inconveniente de ser demasiado amplio, incluyendo helicópteros, globos, drones y otros vehículos aéreos que no son estrictamente aviones. Para mejorar el modelo y caracterizar adecuadamente los aviones, podríamos añadir la frase: "Los aviones son aparatos diseñados por el hombre que vuelan y tienen alas fijas." Esto ayuda a distinguirlos de otros vehículos voladores.

Los modelos están en la base de la ciencia y la ingeniería modernas. Entre todos los tipos de modelos los matemáticos son particularmente poderosos, especialmente los modelos estadísticos, debido a su precisión y capacidad de generalización. Estos modelos nos permiten describir, analizar y predecir una amplia gama de fenómenos, desde el movimiento de los planetas hasta el comportamiento de los mercados financieros y son los que se usan en el aprendizaje automático.

Los modelos matemáticos a menudo incluyen parámetros ajustables que permiten adaptarlos a diferentes situaciones o conjuntos de datos. Es crucial encontrar un equilibrio en la cantidad de parámetros a usar en el modelo. Un modelo con muy pocos parámetros puede ser demasiado simplista y no capturar adecuadamente la complejidad del fenómeno estudiado. Por otro lado, un modelo con demasiados parámetros puede volverse excesivamente complejo, difícil de interpretar y propenso al sobreajuste, donde el modelo se ajusta demasiado a los datos de entrenamiento y pierde capacidad de generalización con lo que no sirve para predecir datos no conocidos. Estos parámetros son los que los algoritmos de aprendizaje automático ajustan a partir de los datos de entrenamiento.

El modelo del sistema solar es un ejemplo excelente de cómo los modelos evolucionan con el tiempo a medida que obtenemos más datos y mejoramos nuestra comprensión. El modelo inicial era geocéntrico, colocando la Tierra en el centro del universo. Sin embargo, este modelo no explicaba adecuadamente las observaciones astronómicas, como el movimiento retrógrado aparente de los planetas.

Copérnico propuso un modelo heliocéntrico, colocando el Sol en el centro. Este modelo explicaba mejor las observaciones, pero aún tenía discrepancias con los datos observados. Kepler mejoró aún más el modelo al proponer que las órbitas de los planetas eran elípticas en lugar de circulares. Este refinamiento (que supone añadir más parámetros, ya que para definir una elipse hacen falta más parámetros que para definir un círculo) permitió explicar casi perfectamente las observaciones astronómicas de la época.

Cada iteración de este modelo del sistema solar representó una mejora significativa en nuestra comprensión del universo, demostrando cómo los modelos evolucionan para adaptarse a nuevas observaciones y conocimientos.

En contraste con el aprendizaje automático, los primeros sistemas de inteligencia artificial utilizaban lo que se denomina inteligencia artificial simbólica, que se fundamenta en la representación explícita del conocimiento y el razonamiento mediante símbolos y reglas lógicas. Estos sistemas, denominados sistemas expertos, y que dominaron el campo de la inteligencia artificial durante las décadas de 1970 y 1980, se basaban en modelos compuestos por conjuntos predefinidos de reglas codificadas manualmente por expertos humanos para resolver problemas en dominios específicos.

Los sistemas expertos tenían varias ventajas:

- Eran transparentes y explicables, ya que las reglas podían ser inspeccionadas y comprendidas por los humanos.
- Funcionaban bien en dominios estrechos y bien definidos.
- No requerían grandes cantidades de datos para su funcionamiento.

Sin embargo, también presentaban limitaciones significativas:

- La creación y mantenimiento de las reglas era un proceso laborioso y propenso a errores.
- Tenían dificultades para manejar situaciones ambiguas o no previstas en sus reglas.
- No se adaptaban fácilmente a nuevos datos o situaciones cambiantes, con lo que era imposible escalar a problemas muy complejos o de dominio abierto

El *aprendizaje automático* surgió como una alternativa que abordaba estas limitaciones. En lugar de depender de reglas predefinidas, los algoritmos de aprendizaje automático pueden:

- Aprender patrones a partir de datos, sin necesidad de programación explícita.
- Adaptarse y mejorar con la experiencia y nuevos datos.
- Manejar problemas complejos y multidimensionales que serían difíciles de abordar con reglas simples.

En realidad, el *aprendizaje automático* es una combinación inteligente de **estadística** (que proporciona los conceptos, métodos y técnicas básicas para crear los modelos y evaluarlos), **investigación operativa** (que proporciona técnicas y métodos de optimización) y **programación** (que proporciona las herramientas para implementar los modelos de forma eficiente).

El resurgimiento del *aprendizaje automático* comenzó en los años 2000, impulsado por:

- El aumento exponencial en la potencia de cómputo disponible.
- La gran cantidad de datos digitales disponibles para entrenar modelos, generados por el auge de Internet y la digitalización masiva.
- Los avances en algoritmos y técnicas, como las redes neuronales profundas.

Y se ha acelerado en los últimos tiempos por el aumento masivo de inversión en el campo por parte de las grandes empresas tecnológicas y los gobiernos.

Hoy en día el *aprendizaje automático* se ha convertido en una herramienta fundamental en diversos campos, desde los sistemas de detección de spam, de concesión de créditos y detección de fraude financiero, o de recomendación de contenidos en plataformas digitales, hasta el reconocimiento de voz, la conducción autónoma, la visión artificial o el diagnóstico médico complejo.

Aunque los sistemas basados en reglas siguen teniendo su lugar en ciertos dominios, el *aprendizaje automático* ha demostrado ser una aproximación

más flexible y potente para abordar una amplia gama de problemas complejos en la era del *big data*.

Pongamos el ejemplo de un sistema informático que pueda distinguir frutas por su aspecto. La forma tradicional sería escribir un complicado código con reglas fijas como "si el objeto es rojo y redondo es una manzana, si es alargado y amarillo es un plátano, etc.", siempre con el inconveniente de que si aparece una nueva fruta hay que codificar una nueva regla. Con el enfoque de *aprendizaje automático* la aproximación es distinta, se alimenta al sistema con miles de fotos de diferentes frutas correctamente etiquetadas y el sistema es capaz de aprender por sí mismo los patrones visuales que distinguen a una fruta de otra y crear un modelo que permita catalogar imágenes que no ha visto anteriormente. ¡Mágico!

En problemas reales, con miles o millones de casos y muchas más variables, esta capacidad de "aprender" automáticamente patrones complejos, que en muchos casos los seres humanos no podemos detectar, se vuelve extraordinariamente potente.

Para que un sistema pueda aprender los patrones correctos, necesitamos alimentarlo con muchos, muchos ejemplos correctamente etiquetados y que estos ejemplos representen de forma equilibrada lo que realmente sucede. Cuantos más ejemplos y más representativos de la realidad sean, mejor será el "aprendizaje". De ahí la importancia de contar con grandes volúmenes de datos de calidad.

Cuando hablamos de *aprendizaje automático* hay que distinguir entre:

- Las **tareas**, que son los problemas específicos que intentamos resolver utilizando el *aprendizaje automático*, como clasificar una nueva instancia entre un conjunto de datos discretos, predecir el valor numérico de una variable, agrupar datos en grupos similares sin tener una etiqueta previa, detectar anomalías en conjuntos de datos, analizar imágenes o generar nuevo contenido.
- Las **técnicas**, que son los métodos y algoritmos específicos para abordar las tareas y que deben ser elegidas cuidadosamente en función de la tarea a realizar y los datos disponibles para obtener el mejor resultado posible.

- Las **herramientas**, que son los recursos de hardware y software que nos permiten implementar las técnicas y resolver las tareas.

TAREAS DE APRENDIZAJE AUTOMÁTICO

Las tareas de *aprendizaje automático* se pueden clasificar como **predictivas** (las orientadas a predecir el valor de uno o varios parámetros no conocidos de una instancia de datos), **descriptivas** (aquellas orientadas a describir un dato o un conjunto de datos, descubriendo grupos, asociaciones y dependencias funcionales o detectando anomalías) y **generativas** (aquellas orientadas a generar nuevos datos, como texto, imágenes, audio o vídeo).

LAS TÉCNICAS DE APRENDIZAJE AUTOMÁTICO

Las técnicas de *aprendizaje automático* son métodos y algoritmos utilizados para entrenar modelos matemáticos que pueden aprender patrones a partir de datos de entrada y ser utilizados posteriormente para realizar las tareas.

Estas técnicas pueden agruparse de diversas maneras, pero una clasificación común es la basada en el tipo de aprendizaje.

1. Aprendizaje supervisado:

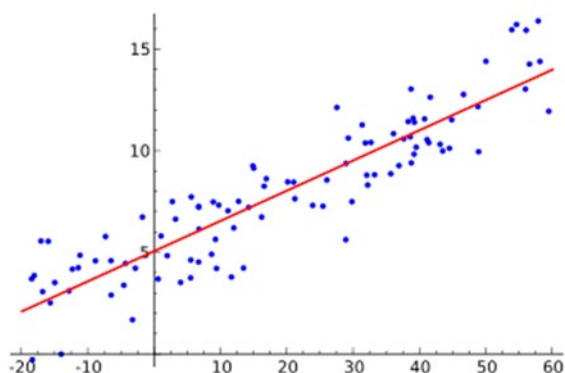
Son técnicas que requieren que los datos de entrada estén etiquetados con respuestas correctas asociadas. El objetivo es desarrollar un modelo que mapee la relación entre las entradas y las salidas para aplicarlo a las nuevas entradas en las que uno o varios parámetros son desconocidos. Por ejemplo, en un problema de clasificación de spam, los datos de entrada serían miles de correos electrónicos etiquetados como "spam" o "no spam". El algoritmo deberá generalizar un modelo capaz de clasificar nuevos correos en una u otra categoría.

Las técnicas de *aprendizaje supervisado* pueden subdividirse a su vez en dos tipos, aquellas en los que el parámetro a predecir solo puede tomar valores de un conjunto discreto de categorías (sí/no, bueno/malo, rojo/verde/azul, grande/mediano/pequeño), que se denominan de **clasificación**, y aquellas en los que el parámetro a predecir es un valor numérico (como, por ejemplo,

el precio de una casa, el tiempo entre fallos de una máquina o el valor de la presión en un determinado proceso), que se denominan de **regresión**.

Un gran número de las tareas de *aprendizaje automático* supervisado de *clasificación* requieren de una salida binaria (o, lo que es lo mismo, que solo puede adoptar dos estados: bueno/malo, spam/no spam, crédito/no crédito), por lo que hay un gran número de técnicas orientadas a este tipo de tareas.

Una de las técnicas más sencillas de regresión es la de **regresión lineal**, modelo matemático heredado de la estadística en el que se pretende hallar una ecuación que permita predecir el valor de una variable numérica (denominada dependiente) a partir de los valores de otras variables numéricas conocidas (denominadas independientes). En el caso de que solo haya una variable independiente (por ejemplo, si se intenta predecir la altura de una persona a partir de su peso), la ecuación a obtener será la de una recta, que se seleccionará para que la distancia entre sus puntos y los de los datos de los que se conoce la variable dependiente e independiente sea la menor posible. Luego se utiliza la ecuación de esa recta para predecir el valor de la variable dependiente para valores nuevos de la variable independiente. En el caso de que haya dos variables independientes la ecuación será la de un plano y, si hay más de dos variables independientes la de un hiperplano, pero al final el resultado es el mismo, una ecuación en la se introducen uno o varios valores de entrada y nos da un valor de salida.



By Sewaqu - Own work,
Public Domain, [https://
commons.wikimedia.org/w/
index.php?curid=11967659](https://commons.wikimedia.org/w/index.php?curid=11967659)

Existen diversas técnicas para clasificación como los *árboles de decisión*, las *máquinas de vectores de soporte* (SVM), el *Naive Bayes*, *K-Nearest Neighbors* o la *regresión logística*, que suelen probarse al realizar un proyecto de aprendizaje automático para seleccionar la que proporciona mejores métricas para los datos de prueba (incluso se prueban combinaciones de ellas con técnicas como el *random forest* o *gradient boosting*).

Para dar una idea de cómo funcionan alguno de estos algoritmos vamos a comentar dos ejemplos sencillos: cómo construir un *árbol de decisión* para clasificar una planta y cómo utilizar *K-Nearest Neighbors* para recomendar una película.

En el primer caso supondremos que, como jardineros, queremos predecir qué tipo de planta florecerá según ciertas características observables. Un *árbol de decisión* puede ayudar a realizar esa predicción basándose en una serie de preguntas sencillas sobre los atributos de la planta.

El proceso comienza construyendo el árbol a partir de los datos de entrenamiento, que en este caso serían un conjunto de plantas cuyo tipo flor ya se conoce. Digamos que tenemos ejemplos de rosas, tulipanes y lirios con datos como altura de la planta, color de las hojas, cantidad de espinas, etc.

Nº DE EJEMPLO	TIPO	ALTURA	ESPINAS	COLOR
1	Rosa	75	Si	Verde
2	Tulipán	40	No	Verde
3	Lirio	65	No	Naranja
4	Rosa	80	Si	Rojizo
5	Tulipán	35	No	Verde
6	Lirio	70	No	Naranja
7	Rosa	60	Si	Verde
8	Tulipán	45	No	Verde
9	Lirio	55	No	Naranja

Cuadro: Datos de entrenamiento para construir un Árbol De Decisión

Para construir el árbol seguimos el siguiente proceso:
En la raíz del árbol se pone una pregunta sobre el atributo más discriminatorio - el que mejor divide los datos en sus clases conocidas (hay técnicas matemáticas para poder determinar cuál es a partir de los datos disponibles).

Por ejemplo: "¿La planta tiene espinas? Sí/No" que nos permite discriminar las rosas

Cada respuesta lleva a un nuevo nodo hijo del árbol, repitiéndose el proceso recursivamente:

Si respuesta es Sí, ¿cuál es ahora la siguiente mejor pregunta para los datos que quedan? Por ejemplo: "¿Hojas rojizas? Sí/No"

Esto continúa rama por rama, eligiendo preguntas para separar los casos que corresponden hasta que en un nodo no haya más preguntas útiles o todos los ejemplos restantes sean de la misma clase. Ahí se marca una planta con la clase de flor predicha.

Una vez construido el árbol, se puede usar para clasificar una nueva planta, comenzando en la raíz y respondiendo la pregunta del nodo según los atributos observados, siguiendo al siguiente nodo dependiendo de la respuesta, y así sucesivamente hasta llegar a un nodo del último nivel del árbol (llamado hoja), que nos dará la clase de flor que el árbol predice para esa planta.



Para explicar el algoritmo de *K-Nearest Neighbors* (K Vecinos Más Cercanos en español) usaremos un recomendador sencillo de películas.

Imagina que tienes una tienda de películas y quieres recomendar películas a tus clientes basándote en sus gustos previos. Usaremos el algoritmo *K-NN* para hacer estas recomendaciones.

Para poder aplicar el algoritmo necesitamos que las películas tengan unas características que nos permitan calcular como de parecidas son (a lo que llamaremos distancia) y una etiqueta sobre el atributo a predecir.

En nuestro caso tendremos como características el nivel de acción (escala del 1 al 10) y el nivel de romance (escala del 1 al 10) y como atributo si le gustaron o no al cliente.

TÍTULO DEL FILM	NIVEL DE ACCIÓN (DEL 1 AL 10)	NIVEL DE ROMANCE (DEL 1 AL 10)	ATRIBUTO (GUSTÓ/NOGUSTÓ)
Rápidos y furiosos	9	2	SI
Diario de una pasión	2	9	NO
Sr. Y sra. Smith	7	6	SI
Titánic	3	8	NO
Misión imposible	8	3	SI

Cuadro: Datos de entrenamiento usando K-Nearest Neighbors (K-NN)

Queremos saber si deberíamos recomendar una nueva película: “Amor y Balas”, con un nivel de acción de 6 y un nivel de romance de 5, representada como (6, 5).

El siguiente paso es encontrar los K vecinos más cercanos (elegimos K=3), lo que significa que buscaremos las 3 películas más cercanas a “Amor y Balas” en nuestro espacio de características calculando su distancia a cada una de las otras películas.

Para calcular la distancia se pueden utilizar distintas fórmulas matemáticas, la más habitual con datos numéricos que tienen la misma importancia (como es nuestro caso, en que le damos el mismo peso al nivel de acción y al de romance) es usar la distancia euclidiana, cuya fórmula para dos dimensiones es $d = \sqrt{[(x_2 - x_1)^2 + (y_2 - y_1)^2]}$, donde x2, y2 son los valores de acción y drama de la película a recomendar y x1, y1 son los de la película a comparar. Con esto tenemos:

TÍTULO DEL FILM	DIMENSIONES	FÓRMULA CÁLCULO DISTANCIA	DISTANCIA
Rápidos y furiosos	(9,2)	$d = \sqrt{[(6 - 9)^2 + (5 - 2)^2]} = \sqrt{[(-3)^2 + 3^2]} = \sqrt{(9 + 9)} = \sqrt{18} \approx 4.24$	4,24
Diario de una pasión	(2,9)	$d = \sqrt{[(6 - 2)^2 + (5 - 9)^2]} = \sqrt{[4^2 + (-4)^2]} = \sqrt{(16 + 16)} = \sqrt{32} \approx 5.66$	5,66
Sr. y Sra. Smith	(7,6)	$d = \sqrt{[(6 - 7)^2 + (5 - 6)^2]} = \sqrt{[(-1)^2 + (-1)^2]} = \sqrt{(1 + 1)} = \sqrt{2} \approx 1.41$	1,41
Titánic	(3,8)	$d = \sqrt{[(6 - 3)^2 + (5 - 8)^2]} = \sqrt{[3^2 + (-3)^2]} = \sqrt{(9 + 9)} = \sqrt{18} \approx 4.24$	4,24
Misión imposible	(8,3)	$d = \sqrt{[(6 - 8)^2 + (5 - 3)^2]} = \sqrt{[(-2)^2 + 2^2]} = \sqrt{(4 + 4)} = \sqrt{8} \approx 2.83$	2,83

Cuadro: Cálculo de la distancia euclidiana para dos dimensiones

Ordenando estas distancias de menor a mayor:

"Sr. y Sra. Smith": 1.41 que Le gustó

"Misión Imposible": 2.83 que Le gustó

"Rápidos y Furiosos" / "Titanic": 4.24 cuyos valores son Le gustó/ No le gustó

"Diario de una pasión": 5.66 que No le gustó.

Por lo tanto, los 3 vecinos más cercanos son "Sr. y Sra. Smith", "Misión Imposible" y "Rápidos y Furiosos" (o "Titanic", dependiendo de cómo maneje los empates).

Con eso tenemos que, para los 3 vecinos más cercanos hay o 3 "Le gustó" o 2 "Le gustó" y un "No le gustó" (depende de si elegimos "Rápidos y Furiosos" o "Titanic"), con lo que nuestro algoritmo predice que a este cliente también le gustará "Amor y Balas" y podemos recomendársela.

2. Aprendizaje no supervisado:

No se proporcionan salidas o etiquetas en los datos de entrenamiento a los algoritmos, que deben encontrar patrones interesantes, irregularidades o relacionar por sí mismos los datos de entrada. Un caso típico es la agrupación o *clustering*, donde el sistema agrupa automáticamente los datos en grupos con características similares. Por ejemplo, un sistema de marketing no supervisado podría detectar automáticamente segmentos de clientes con comportamientos de compra similares a partir de sus datos transaccionales.

Algunos ejemplos son el *K-means* o el *Clustering jerárquico* para el *agrupamiento*, *Apriori* para reglas de asociación o *Isolation Forest* para *detección de anomalías*.

3. Aprendizaje Semi-supervisado:

Se utilizan datos etiquetados y no etiquetados para mejorar el rendimiento de los modelos.

4. Aprendizaje por refuerzo:

Un agente aprende a tomar decisiones interactuando con un entorno y recibiendo retroalimentación en forma de recompensas o penalizaciones. El

objetivo es aprender una política que maximice la recompensa total a largo plazo. Se utiliza para la programación de robots o juegos.

5. Aprendizaje profundo:

Se utilizan redes neuronales artificiales con múltiples capas de abstracción que aprenden distintas características de forma jerárquica (de ahí el nombre de profundo).

Las redes neuronales, inspiradas en cómo funciona el cerebro humano, están formadas un gran número de ecuaciones matemáticas encadenadas que representan una red de nodos conectados entre sí, donde cada conexión tiene un "*peso*" que determina cuánto influye en el resultado final. La red va ajustando estos pesos a medida que aprende de los datos, hasta que consigue hacer predicciones precisas.

Hay una *capa de entrada*, un número de *capas ocultas* y una *capa de salida*, cada una con un número de *neuronas*.

La forma más común de *aprendizaje profundo* es el *aprendizaje supervisado*, aunque hay técnicas de entrenamiento no supervisado y *semi-supervisado* e incluso de *aprendizaje por refuerzo* para entrenar *redes neuronales profundas*.

Las *redes neuronales de aprendizaje profundo*, con miles de millones de neuronas distribuidas en muchas capas, están detrás de los últimos avances en *procesamiento del lenguaje natural* (traducción automática y generación de texto), *procesamiento y síntesis del habla*, *visión artificial*, *diagnóstico médico o conducción autónoma*.

6. Aprendizaje por transferencia:

Es una técnica utilizada en *aprendizaje profundo* en la que un modelo desarrollado para una tarea se reutiliza como punto de partida para un modelo a utilizar en una segunda tarea relacionada.

LAS HERRAMIENTAS DE APRENDIZAJE AUTOMÁTICO

Las herramientas nos permiten limpiar y cargar los datos, segmentarlos y aplicar y evaluar diversas técnicas para resolver una tarea, facilitando la selección de la mejor opción en cada caso.

Hay librerías para lenguajes de programación como Python y R (Sci-kit-learn, TensorFlow, Pytorch, OpenCV, Prophet), paquetes informáticos específicos de aprendizaje automático, muchos de ellos con interfaz gráfica (como Orange, Knime, Rapid Miner o Watson Studio de IBM) y hardware específico (GPU, TPU) para acelerar estos procesos.

EL PROCESO DE APRENDIZAJE AUTOMÁTICO

Pero ¿cuál es el proceso a seguir para crear y utilizar un modelo de *aprendizaje automático*?

Hay varios modelos que tratan de sintetizar los pasos a llevar a cabo, siendo uno de los más populares el denominado CRISP-DM, que es un modelo estándar abierto.

Las etapas que define son: Compresión de negocio $\rightleftharpoons$ comprensión de los datos $\rightarrow$ preparación de los datos $\rightleftharpoons$ creación del modelo $\rightarrow$ Evaluación $\rightarrow$ Despliegue



By Kenneth Jensen -
Own work based on:
<ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/18.0/en/ModelerCRISPDm.pdf>
(Figure 1), CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=24930610>

Se empieza por definir la tarea a realizar, su tipo y qué datos necesitamos para llevarla a cabo, analizando primero el negocio y luego los datos disponibles. En muchos casos estos pasos requieren procesos iterativos de entendimiento del negocio y los datos.

Luego tenemos que buscar los datos, que suelen provenir de fuentes diversas (ficheros, bases de datos propias de la empresa, sistemas externos,

webs, entre otros), tener errores y no estar en el formato necesario para entrenar el modelo, con lo que hay que llevar a cabo un proceso denominado **ETL** por sus siglas en inglés para *extracción, transformación y carga* (load).

En el proceso **ETL** se limpiarán los datos (corrigiendo valores erróneos, eliminando duplicados, identificando datos faltantes y decidiendo qué hacer con ellos), se adaptará su formato (maneja los datos desbalanceados, codificando variables categóricas o creando nuevas características si es conveniente) y se consolidarán en un solo fichero con el que alimentar el proceso de aprendizaje automático. Este proceso suele ser muy laborioso y es el que más tiempo consume normalmente en los proyectos de *aprendizaje automático*, con entre un 60 y un 80% del tiempo total del proyecto en la mayoría de los casos.

Un proceso **ETL** bien ejecutado puede ser la diferencia entre un modelo mediocre y uno excelente.

Con la maduración de los proyectos se suelen ir consiguiendo procesos más eficientes, aunque, en muchos casos, este proceso puede llegar a ser continuo en el tiempo por la evolución de los datos de entrada.

Merece la pena reseñar la importancia de que los datos estén balanceados para el propósito del modelo. Esto se refiere a que todas las categorías están representadas de manera aproximadamente igual, ya que, cuando los datos están desbalanceados, puede llevar a modelos sesgados que favorecen la clase mayoritaria.

Un ejemplo claro de lo anterior sería un modelo para detectar fraudes en transacciones de tarjetas de crédito en el que, como datos de entrenamiento, de un total de un millón de transacciones tenemos 1.000 fraudulentas (0,1%) y el resto legítimas. Este modelo podría identificar todas las transacciones como legítimas y conseguiría una precisión del 99,9%, pero fallaría completamente en detectar fraudes reales, que es su objetivo principal.

Hay estrategias para balancear los datos, dando más peso a los casos de la clase minoritaria o menos a los de la mayoritaria, pero hay que ser conscientes del problema al cargar los datos y aplicar estas técnicas correctamente.

En las técnicas de aprendizaje supervisado, una vez cargados los datos debemos segmentarlos, seleccionando una parte para entrenar el modelo o modelos y guardando otra parte para probar los modelos que generemos,

ya que probar los modelos con los mismos datos con los que los hemos entrenados nos llevaría a un problema conocido como **sobreajuste** (*overfitting* en inglés), en el que el modelo se ajusta tanto a los datos de entrenamiento que proporciona unas métricas excelentes cuando se prueba con ellos, pero no captura los patrones generales subyacentes en los datos, con lo que su rendimiento con datos nuevos es significativamente peor y, por lo tanto, no sirve.

Normalmente se suelen hacer tres conjuntos de datos: **entrenamiento**, **validación** (para poder comparar entre varios modelos y realizar ajustes) y **prueba** para la evaluación final del modelo seleccionado.

Hay técnicas para reutilizar los mismos datos si se tienen pocos, dividiéndolos y utilizando los subconjuntos de forma inteligente, de forma que nunca se entrene y pruebe un modelo con los mismos datos. En los casos como las *redes neuronales profundas*, que requieren conjuntos ingentes de datos, hay técnicas de generación de datos sintéticos para completar su entrenamiento y evaluación.

Con los conjuntos de entrenamiento, validación y prueba definidos, el siguiente paso es entrenar diversos modelos usando una o varias de las técnicas específicas adaptadas a la tarea y probando con distintos valores para los parámetros de cada técnica, de forma que acabemos seleccionando el modelo y los parámetros que nos den las mejores métricas de calidad. Este paso se realiza con programas informáticos que lanzan de forma automática el entrenamiento de distintos modelos con distintos parámetros para poder comparar su resultado, y puede requerir tener que modificar los datos disponibles una vez obtenidas las primeras conclusiones, con lo que también se produce una interacción con el paso anterior para optimizar los resultados obtenidos.

A la hora de evaluar la calidad de los resultados de un *modelo de clasificación* se utilizan diferentes cocientes entre la cantidad de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos de cada categoría, ya que cada uno de estos cocientes nos da la calidad de un modelo para una tarea específica. Los más comunes para clasificación son:

- **Matriz de confusión:** Tabla en la que se muestran los verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos para cada clase, y que nos da una idea del rendimiento del modelo para cada categoría.
- **Tasa de acierto (Accuracy):** (Verdaderos positivos+Verdaderos negativos) / Total, que puede ser engañosa en casos donde los datos están desbalanceados, como hemos visto.
- **Precisión (Precision):** Verdaderos positivos/(Verdaderos positivos+Falsos positivos) que puede ser representativa cuando el coste de los falsos positivos es alto.
- **Sensibilidad o recuerdo (Recall):** Verdaderos Positivos / (Verdaderos Positivos + Falsos Negativos), importante cuando el costo de los falsos negativos es alto
- **F1 score:** $2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$ que nos da un balance entre la precisión y la sensibilidad.

Para regresión se suelen usar los errores medios entre los valores reales y los proporcionados por la ecuación (**error cuadrático medio** que penaliza los errores grandes frente a los pequeños, raíz del error anterior que nos da el error en las mismas unidades que la variable a predecir o un parámetro denominado R^2 que indica lo bien que el modelo explica la variabilidad de los datos y que es mejor cuanto más se acerca a 1).

Hay también métricas específicas para evaluar los modelos más avanzados de *aprendizaje profundo*, como el índice de Jaccard para la segmentación de imágenes o *Perplexity* para el procesamiento del lenguaje natural.

El último paso, una vez seleccionado el mejor modelo, es el despliegue de la aplicación para su uso en el mundo real.

LOS DESAFÍOS DEL APRENDIZAJE AUTOMÁTICO

El *aprendizaje automático*, y en particular el *aprendizaje profundo*, han experimentado avances a una velocidad de vértigo, habiendo conseguido en los últimos años logros que hasta hace poco parecían imposibles, pero se enfrenta a desafíos importantes que deben ser resueltos para continuar avanzando. Haremos un resumen de los más importantes, empezando por los *desafíos técnicos*:

- ***Necesidad de cantidades de datos cada vez más grandes***, datos que deben conseguirse respetando las regulaciones de tratamiento de datos personales y deben tratarse adecuadamente, con el costo asociado.
- ***Posible conflicto con las leyes de protección de los derechos de autor***, lo que puede complicar todavía más la disponibilidad de datos de entrenamiento.
- ***Dificultad para entender y explicar las decisiones del modelo*** en algunas técnicas como modelos de *aprendizaje profundo*, lo que impide su uso en aplicaciones en campos sensibles como medicina o finanzas.
- ***Problemas de sesgo y equidad***, pues los modelos se entrenan con los datos disponibles en internet y reproducen y amplifican sus sesgos sociales y estereotipos, puede llevar a decisiones injustas en áreas como contratación, justicia penal, y servicios financieros (por ejemplo, se han detectado problemas en algunos sistemas de reconocimiento facial para mujeres de piel oscura, al estar dominados los conjuntos de datos usados para entrenarlos por imágenes de hombres de piel blanca).
- ***Tendencia a la invención de datos*** en los modelos de *aprendizaje profundo* (conocida como alucinación), que hace poco confiable sus respuestas e implican la necesidad de supervisión humana.
- ***Problemas de seguridad y privacidad***: Los modelos de aprendizaje profundo son sistemas muy complejos que trabajan con grandes cantidades de datos, lo que los hace vulnerables a ataques y fallos de seguridad que, dada su potencia creciente, podrían ser potencialmente muy peligrosos.
- ***Problemas de reproducibilidad y estabilidad***: Los modelos generados tienen dificultad para reproducir resultados debido a la naturaleza estocástica del entrenamiento, además pueden fallar inesperadamente ante datos ligeramente diferentes y pueden producir soluciones muy diferentes ante la misma pregunta, algunas correctas y otras erróneas.
- ***Problemas de eficiencia computacional y energética***: El entrenamiento de modelos grandes de aprendizaje profundo requiere recursos computacionales significativos que son caros y consumen mucha energía, teniendo una huella de carbono que crece a gran velocidad
- ***Problemas con el aprendizaje continuo y la generalización***: Los modelos actuales tienen dificultades para aprender continuamente y para generalizar el conocimiento o transferirlo de un dominio a otro, estando lejos

todavía de la inteligencia artificial general (aquella con capacidades generales e ilimitadas que puede adaptarse a cualquier ámbito o dominio transfiriendo conocimiento que ya tenga de otros dominios).

Pero también se plantean otros *problemas de índole estrictamente ético*:

- **Autonomía y Deshumanización:** La delegación de decisiones importantes a las máquinas puede reducir el control humano y la responsabilidad moral.
- **Desplazamiento Laboral:** La automatización impulsada por el aprendizaje automático puede llevar a la pérdida de empleos, afectando a la economía y a los trabajadores, aunque, como en cualquier innovación, también se crearán nuevos trabajos que generarán empleo.
- **Responsabilidad Legal:** Determinar quién es responsable cuando un sistema de inteligencia artificial causa daño es complejo.
- **Manipulación y Desinformación:** Los modelos pueden ser utilizados para crear contenido falso o manipular opiniones públicas, como en el caso de los deepfakes.

Abordar estos problemas, además de los avances técnicos, requiere un enfoque multidisciplinar que incluya regulaciones, prácticas de desarrollo ético y una mayor conciencia pública.

CONCLUSIONES

El *aprendizaje automático* ha revolucionado la forma en que abordamos problemas complejos y procesamos grandes cantidades de datos, convirtiéndose en la base de los últimos avances en inteligencia artificial, que nos ha permitidos logros como el reconocimiento del habla, la visión artificial o la conducción autónoma. Combinando estadística, investigación operativa y programación, el *aprendizaje automático* permite crear modelos que aprenden patrones de los datos sin necesidad de programación explícita, adaptándose y mejorando con nuevos datos.

Con el *aprendizaje automático* podemos predecir parámetros, descubrir relaciones en conjuntos de datos y generar nuevo conocimiento multimedia.

Sin embargo, a pesar de sus enormes avances y potencial transformador, el *aprendizaje automático* aún enfrenta importantes desafíos técnicos y éticos que será clave abordar con enfoques multidisciplinares, siendo fundamental establecer un sólido marco ético, legal y de gobernanza que garantice que esta poderosa tecnología se aplica de manera responsable y alineada con los valores humanos fundamentales.

Para la redacción de este capítulo se han utilizado diversos modelos de inteligencia artificial generativa como herramienta de apoyo. 

EL DEEP LEARNING EN SISTEMAS DE CCTV

🗨 Ramón Segarra

El campo de la videovigilancia (Circuito Cerrado de Televisión, CCTV) ha experimentado una revolución significativa con la introducción del Deep Learning (aprendizaje profundo) en cámaras IP y grabadores de red (NVR). Esta tecnología de inteligencia artificial ha transformado la concepción del uso de las cámaras y ha ampliado de manera espectacular el número de aplicaciones en las que se hace uso de las cámaras dotadas de esta tecnología.

El Deep Learning es una rama avanzada del machine learning, utiliza redes neuronales artificiales para imitar el funcionamiento del cerebro humano. En el contexto de los sistemas CCTV, esta tecnología permite a las cámaras IP y NVRs procesar y analizar imágenes de video de forma autónoma, identificando objetos, personas y comportamientos con gran precisión, en tiempo real¹.

FUNDAMENTOS DEL DEEP LEARNING (APRENDIZAJE PROFUNDO) EN CCTV

Arquitectura de Redes Neuronales

El núcleo del aprendizaje profundo en sistemas CCTV son las redes neuronales convolucionales (CNN). Estas redes están diseñadas específicamente para procesar datos visuales y consisten en múltiples capas de neuronas artificiales interconectadas. Cada capa extrae características cada vez más complejas de la imagen, desde bordes y texturas simples en las primeras capas hasta formas y objetos completos en las capas más profundas².

Proceso de Entrenamiento

El entrenamiento de estos sistemas implica exponer la red neuronal a grandes cantidades de datos etiquetados. En el caso de los sistemas CCTV, esto significa

1 Fuente: <https://www.visiotechsecurity.com/es/noticias/365-camaras-de-seguridad-con-inteligencia-artificial> (05/09/2024)

2 Fuente: https://revistainnovacion.com/nota/10337/como_deep_learning_beneficia_el_sector_de_la_seguridad/ (05/09/2024)

alimentar la red con millones de imágenes de personas, vehículos, objetos y situaciones diversas. A través de este proceso, la red aprende a reconocer patrones y características distintivas, mejorando continuamente su precisión³.

Inferencia en Tiempo Real

Una vez entrenada, la red neuronal puede realizar inferencias en tiempo real sobre nuevas imágenes o secuencias de video. Esto permite a las cámaras IP y a los NVRs equipados con esta tecnología, analizar el flujo de video en vivo, detectando y clasificando objetos, personas y eventos de interés de forma instantánea⁴.

APLICACIONES DEL DEEP LEARNING EN CÁMARAS CCTV

Detección y Reconocimiento Facial

Una de las aplicaciones más impactantes, y controvertidas, de esta tecnología en cámaras CCTV es la detección y reconocimiento facial. Las cámaras equipadas con esta tecnología realizan el siguiente proceso:

- Detectan rostros en tiempo real en un flujo de video; incluso si el rostro está girado, el individuo porta gafas de sol, lleva peluca o cualquier objeto que a modo de sombrero cubra la cabeza y parte de la cara.
- Extraen características faciales únicas (sexo, franja de edad, uso de gafas, uso de mascarilla, color de pelo, ...)
- Comparan estas características con una base de datos de rostros conocidos, que se ha alimentado previamente. Usualmente se pueden rellenar varias librerías de centenares de miles de rostros, pudiendo establecer diferentes acciones en función de la librería en la que se encuentre el rostro identificado.
- Identificar individuos específicos con un alto grado de precisión, proporcionando el dato del tanto por ciento de coincidencia⁵.

3 Fuente: <https://www.visiotechsecurity.com/es/noticias/365-camaras-de-seguridad-con-inteligencia-artificial> (05/09/2024)

4 Fuente: <https://www.youtube.com/watch?v=n9N3B4x61AM> (05/09/2024)

5 Fuente: <https://www.visiotechsecurity.com/es/noticias/365-camaras-de-seguridad-con-inteligencia-artificial> (05/09/2024)

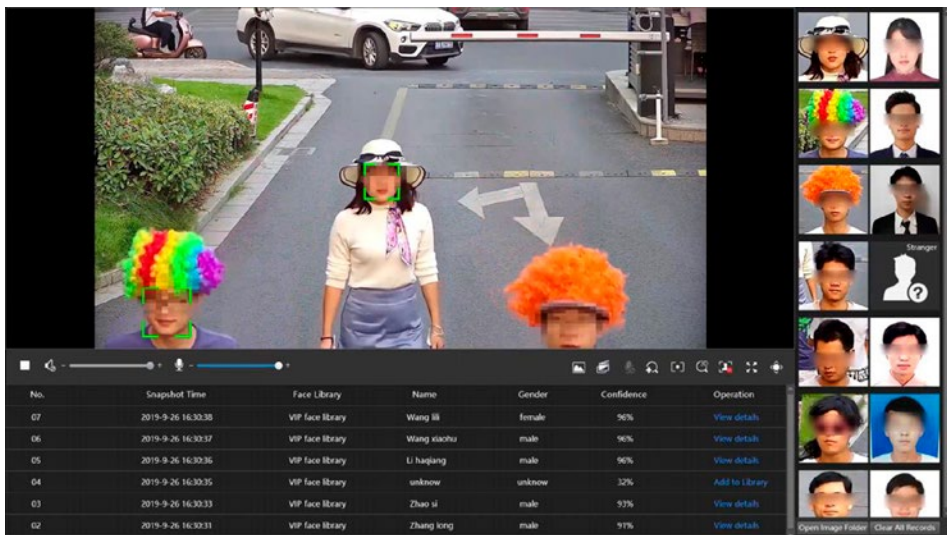


Figura 1: Reconocimiento facial metadatos en cámaras de CCTV dotadas de Deep Learning.

En la UE el uso de esta tecnología está muy restringido y sólo puede hacerse uso, con el consentimiento escrito, explícito y expreso de las personas sobre las que se realiza el reconocimiento facial. Si bien es cierto que, en países con una legislación más laxa, esta capacidad tiene aplicaciones en seguridad, control de acceso, identificación de ladrones reincidentes y análisis de comportamiento del cliente en entornos comerciales.

Detección y Seguimiento de Objetos

Las cámaras con esta tecnología de IA pueden detectar, identificar y seguir (lo que se conoce como autotracking) objetos específicos en una escena, como:

- Vehículos de cuatro ruedas. En este caso se identifica marca, modelo, color, dirección de paso e incluso la realización de giros o acciones no permitidas en la vía.
- Vehículos de dos ruedas. En este caso se identifica las características de la persona que la conduce.
- Personas, pudiendo determinar su sexo, color de pelo, franja de edad, uso de gafas, uso de mascarilla, color de la ropa (superior e inferior), uso de mochila,

- Animales
- Objetos abandonados (no estaban en la escena y aparecen a partir de un determinado momento), objetos sustraídos (estaban en la escena y desaparecen a partir de un determinado momento) y objetos bloqueantes (en un determinado momento aparecen en la escena y bloquean una vía de paso).

Esta función es particularmente útil en la gestión del tráfico, la seguridad perimetral y la prevención de robos⁶.

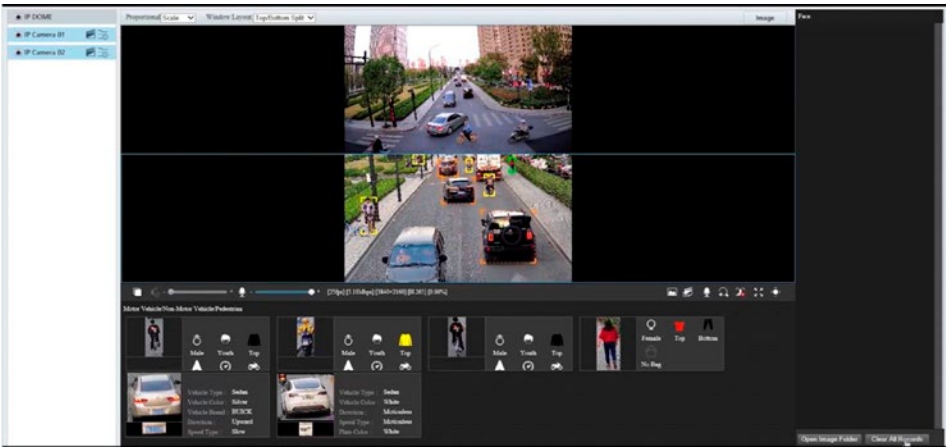


Figura 2: Detección de objetos, identificación y extracción de metadatos en cámaras de CCTV dotadas de Deep Learning.

Análisis de Comportamiento

El aprendizaje profundo permite a las cámaras CCTV analizar patrones de movimiento y comportamiento, identificando situaciones anómalas o potencialmente peligrosas, como:

- Merodeo sospechoso.
- Peleas.
- Caídas.
- Uso del móvil (donde no esté permitido).
- Presencia de fuego.
- Presencia de humo.

6 Fuente: <https://www.altaico.es/camaras-de-vigilancia-con-inteligencia-artificial/> (05/09/2024)

- No uso de casco (en lugares en los que su uso es obligatorio).
- Persona dormida.
- Ausencia (en caso de que ésta no pueda darse).
- Aparcamiento ilegal (en una zona no permitida).
- Uso de la ropa laboral exigida (en los casos que así se exija).
- Movimientos erráticos.
- Aglomeraciones inusuales.
- Accesos a áreas restringidas.

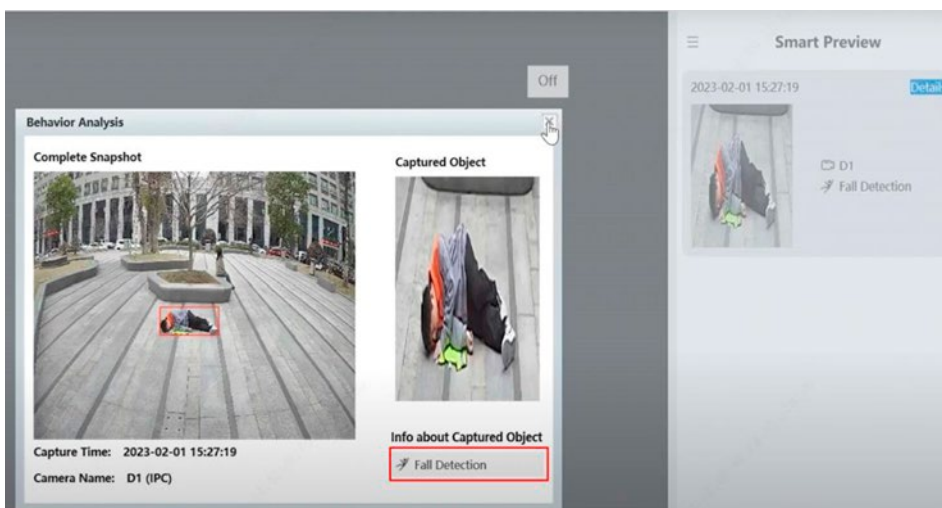


Figura 3: Detección de comportamientos (caída en el ejemplo) con equipos de CCTV con Deep Learning

Esta capacidad abre la puerta a un gran número de aplicaciones en la prevención de riesgos laborales, la prevención de delitos, la gestión de tráfico, la gestión multitudes y seguridad ciudadana⁷.

Conteo de Personas y Análisis de Multitudes

Las cámaras inteligentes pueden realizar un conteo preciso de personas en un área determinada y analizar el flujo de multitudes. Esto es útil para:

⁷ Fuente: <https://www.youtube.com/watch?v=n9N3B4x61AM> (05/09/2024)

- Gestión de colas.
- Optimización de espacios comerciales (mediante mapas de calor).
- Control de aforo en eventos.
- Análisis de patrones de tráfico peatonal (mediante mapas de calor)⁸.

DEEP LEARNING EN GRABADORES DE RED (NVR)

Procesamiento Centralizado

Los NVRs equipados con esta tecnología, actúan como centros de procesamiento para múltiples cámaras. Esto permite:

- Análisis de video más potente y eficiente (se distribuyen mejor los recursos).
- Correlación de datos de múltiples fuentes (las diferentes cámaras).
- Reducción de la carga de procesamiento en las cámaras individuales⁹.

Búsqueda Inteligente de Video

Una de las características más potentes de los NVRs con inteligencia artificial es la capacidad de realizar búsquedas inteligentes en grandes volúmenes de datos grabados (son necesarios GB¹⁰ e incluso TB¹¹). Los usuarios pueden buscar:

- Personas específicas basadas en características físicas o vestimenta.
- Enmarcando a la persona o vehículo que se quiere buscar.
- Vehículos por tipo, color o número de placa.
- Eventos específicos como la aparición de un objeto o un comportamiento particular¹².

8 Fuente: <https://www.visiotechsecurity.com/es/noticias/365-camaras-de-seguridad-con-inteligencia-artificial> (05/09/2024)

9 Fuente: https://revistainnovacion.com/nota/10337/como_deep_learning_beneficia_el_sector_de_la_seguridad / (05/09/2024)

10 GB: Gigabyte (1.024 MegaBytes)

11 TB: Terabyte (1.024 Gigabyte=1.024 x 1.024 Megabytes)

12 Fuente: <https://www.youtube.com/watch?v=ngN3B4x61AM> (05/09/2024)

Esta función reduce drásticamente el tiempo necesario para encontrar evidencia relevante en investigaciones.

Análisis Forense Avanzado

Los NVRs con esta tecnología de Inteligencia artificial ofrecen capacidades de análisis forense que van más allá de la simple reproducción de video. Pueden:

- Reconstruir la trayectoria de individuos o vehículos a través de múltiples cámaras.
- Identificar patrones de comportamiento a lo largo del tiempo.
- Generar informes detallados sobre incidentes o tendencias¹³.

VANTAJAS DEL DEEP LEARNING EN SISTEMAS CCTV

Mayor Precisión

Comparado con los sistemas de análisis de video tradicionales, el Deep Learning ofrece una precisión significativamente mayor en la detección y clasificación de objetos y eventos. Esto se traduce en:

- Reducción dramática de las falsas alarmas.
- Mayor confiabilidad en la detección de incidentes reales.
- Mejor capacidad para distinguir entre comportamientos normales y anómalos¹⁴.

Adaptabilidad a Diferentes Entornos

Los sistemas basados en esta tecnología tienen una notable capacidad para adaptarse a diferentes entornos y condiciones. Pueden:

- Funcionar eficazmente en diversas condiciones de iluminación.
- Ajustarse a cambios en el fondo o la disposición de la escena.

13 Fuente: <https://www.altaico.es/camaras-de-vigilancia-con-inteligencia-artificial/> (05/09/2024)

14 Fuente: <https://www.visiotechsecurity.com/es/noticias/365-camaras-de-seguridad-con-inteligencia-artificial> (05/09/2024)

- Manejar oclusiones parciales y ángulos de visión complejos¹⁵.

Escalabilidad

La naturaleza de este tipo de inteligencia artificial permite que estos sistemas mejoren continuamente con más datos y entrenamiento. Esto significa que:

- El rendimiento del sistema puede mejorar con el tiempo.
- Se pueden añadir nuevas capacidades de detección sin necesidad de reemplazar el hardware, a través de la actualización con nuevos firmwares¹⁶ (actualización del software de los equipos)¹⁷.

Automatización de Tareas de Seguridad

Así mismo, permite automatizar muchas tareas que tradicionalmente requerían intervención humana constante, como:

- Detección automática de eventos y comportamientos, marcados como de interés.
- Monitorización simultánea y continua de un gran número de cámaras.
- Detección inmediata de intrusiones o comportamientos sospechosos.
- Generación automática de alertas e informes¹⁸.

DESAFÍOS Y CONSIDERACIONES

Requisitos de Hardware

La implementación del aprendizaje profundo en sistemas CCTV requiere hardware especializado, incluyendo:

15 Fuente: https://revistainnovacion.com/nota/10337/como_deep_learning_beneficia_el_sector_de_la_seguridad/ (05/09/2024)

16 Firmware: El firmware es el programa que controla los circuitos electrónicos de un dispositivo y que se almacena en una memoria ROM. Fuente: <https://www.xataka.com/basics/que-firmware-que-se-diferencia-drivers> (03/09/2024)

17 Fuente: <https://www.youtube.com/watch?v=ngN3B4x61AM> (05/09/2024)

18 Fuente: <https://www.altaico.es/camaras-de-vigilancia-con-inteligencia-artificial/> (05/09/2024)

- Cámaras con procesadores de alto rendimiento (GPU^{19s}).
- NVRs con GPUs potentes para el procesamiento de video y capacidad de almacenamiento (Terabytes).
- Infraestructura de red robusta para manejar el aumento en el flujo de datos (si bien estos datos suponen una pequeña fracción del ancho de banda requerido por el video transmitido por las cámaras)²⁰.

Privacidad y Consideraciones Éticas

El uso de tecnologías de reconocimiento facial y análisis de comportamiento plantea importantes cuestiones éticas y de privacidad. Es crucial:

- Cumplir con las regulaciones de protección de datos.
- Implementar medidas de seguridad robustas para proteger la información personal.
- Utilizar técnicas de pixelado automático de matrículas, rostros o partes del cuerpo.
- Establecer mecanismos de cifrado de las transmisiones de video, mediante el uso de certificados electrónicos reconocidos o contraseñas fuertes. En esta línea, en enero de 2024, ONVIF²¹ (Open Network Video

19 GPU: La GPU es un procesador formado por muchos núcleos más pequeños y especializados. Al trabajar juntos, los núcleos aumentan el desempeño de forma considerable cuando una tarea de procesamiento puede dividirse entre varios núcleos al mismo tiempo (o en paralelo). La GPU es un componente esencial del gaming moderno, ya que permite obtener imágenes de mejor calidad y una mayor fluidez en el juego. Las GPUs también son útiles en la IA. Fuente: www.intel.la (03/09/2024)

20 Fuente: https://revistainnovacion.com/nota/10337/como_deep_learning_beneficia_el_sector_de_la_seguridad/ (05/09/2024)

21 ONVIF: Es una organización sin ánimo de lucro que tiene como objetivo desarrollar estándares abiertos para la interoperabilidad de dispositivos de videovigilancia IP. Fundada en 2008, ONVIF ha desempeñado un papel crucial en la creación de normas que permiten a los diferentes fabricantes de equipos de videovigilancia trabajar juntos de manera más eficiente. Fuente: Blog de Ramón Segarra Clara. Add-on TLS de ONVIF. URL: <https://ramon-segarraclara.com/add-on-tls-de-onvif> (03/09/2024)

Interface Forum) lanzo el add-on²² TLS²³ para estos fines de forma transversal a cualquier fabricante, adscrito a ONVIF.

- Establecer políticas claras sobre el uso y almacenamiento de datos biométricos²⁴.

Necesidad de Entrenamiento Continuo

Para mantener y mejorar la precisión del sistema, es necesario que los fabricantes:

- Actualizar regularmente los modelos de Deep Learning con nuevos datos.
- Adaptar los modelos a cambios en el entorno o nuevos requisitos de seguridad.
- Deben mantener un equipo técnico capacitado para gestionar y optimizar el sistema²⁵.

IMPLEMENTACIÓN Y MEJORES PRÁCTICAS

Evaluación de Necesidades

Antes de implementar un sistema CCTV con inteligencia artificial, es crucial:

22 Add-on: Los add-on de ONVIF son extensiones opcionales que permiten a los fabricantes agregar funciones específicas a sus dispositivos. Estas extensiones pueden incluir características avanzadas como el control PTZ (Pan-Tilt-Zoom), detección de movimiento, configuraciones de privacidad y más. La importancia de los add-on radica en su capacidad para mejorar la versatilidad y la funcionalidad de los dispositivos de videovigilancia, permitiendo a los usuarios adaptar sus sistemas a sus necesidades específicas. Fuente: Blog de Ramón Segarra Clara. Add-on TLS de ONVIF. URL: <https://ramonsegarraclara.com/add-on-tls-de-onvif> (03/09/2024)

23 TLS: Transport Layer Security, o TLS, es un protocolo de seguridad ampliamente adoptado, diseñado para facilitar la privacidad y la seguridad de los datos en las comunicaciones por Internet. Un caso de uso primario de TLS es la encriptación de las comunicaciones entre aplicaciones web y servidores, como los navegadores que cargan un sitio web. Fuente: <https://www.cloudflare.com/es-es/learning/ssl/transport-layer-security-tls/> (03/09/2024)

24 Fuente: <https://www.visiotechsecurity.com/es/noticias/365-camaras-de-seguridad-con-inteligencia-artificial> (05/09/2024)

25 Fuente: <https://www.youtube.com/watch?v=ngN3B4x61AM> (05/09/2024)

- Identificar claramente las necesidades y los objetivos de seguridad y operativos.
- Evaluar el entorno físico y las condiciones de iluminación.
- Determinar los tipos de eventos o comportamientos que es necesario detectar²⁶.
- Realizar pruebas de campo y analizar los resultados obtenidos.

Selección de Hardware

La elección del hardware adecuado es fundamental:

- Seleccionar cámaras y/o NVR que realicen los análisis que son requisitos del proyecto.
- Optar por cámaras con chipset de alta calidad, con la suficiente potencia y que garanticen la ciberseguridad del equipo. Actualmente hay una gran controversia en este sentido debido a la regulación establecida en EE.UU con el NDAA²⁷
- Seleccionar NVRs con capacidad de procesamiento suficiente (GPUs) para manejar el análisis del número de cámaras que se le van a conectar, siempre que las cámaras no cuenten con esta tecnología.
- Asegurarse de que la infraestructura de red pueda soportar el ancho de banda exigido por el sistema²⁸.

Configuración y Calibración

Una configuración cuidadosa es esencial para maximizar el rendimiento:

26 Fuente: <https://www.altaico.es/camaras-de-vigilancia-con-inteligencia-artificial/> (05/09/2024)

27 NDAA: En CCTV el cumplimiento de la NDAA se refiere a cumplir con los requisitos de la Ley de Autorización de Defensa Nacional (NDAA), específicamente la Sección 889 de la Ley de 2019, que restringe la adquisición de equipos de videovigilancia de ciertas empresas chinas, que utilizan chipsets o tecnologías consideradas como no seguras.

28 Fuente: https://revistainnovacion.com/nota/10337/como_deep_learning_beneficia_el_sector_de_la_seguridad/ (05/09/2024)

- Posicionar las cámaras según las indicaciones del fabricante en cuanto a distancias, alturas máximas de instalación y ángulos de inclinación con respecto a la horizontal y/o la vertical. Además, la instalación debe realizarse con el fin de que los objetos a detectar/analizar deben contar con al menos un tamaño mínimo en píxeles²⁹ para una detección adecuada.
- Ajustar el área o áreas de detección que son de interés, evitando que haya objetos en las mismas tras los cuales las cámaras no puedan realizar correctamente su trabajo de análisis y/o detección.
- Calibrar los algoritmos de detección para minimizar falsos positivos (falsas alarmas) y los falsos negativos, que son si cabe más críticos.
- Ajustar la sensibilidad de los sistemas de alerta según las necesidades específicas del entorno, y la presencia o no de objetos, animales en su mayoría, que puedan producir falsas alarmas³⁰.

Integración con Sistemas Existentes

Para obtener el máximo beneficio, es importante:

- Que los sistemas dispongan de capacidad de integración, para lo cual es necesario que cumplan con el Perfil M de ONVIF³¹. Este perfil es efectivo únicamente cuando las dos partes a integrar lo cumplen.
- Establecer acciones claras a cada uno de los eventos/alertas generadas por el sistema.

29 Píxel: Un píxel es la parte de un sensor que recoge fotones para convertirlos en fotoelectrones. Varios píxeles cubren la superficie del sensor para que se pueda determinar tanto la cantidad de fotones detectados como la ubicación de estos fotones. Las cámaras miden su resolución en píxeles, y más frecuentemente en millones de píxeles (Megapíxeles).

30 Fuente: <https://www.visiotechsecurity.com/es/noticias/365-camaras-de-seguridad-con-inteligencia-artificial> (05/09/2024)

31 Perfil M de ONVIF: El Perfil M de ONVIF se centra en el intercambio de metadatos y eventos para aplicaciones de análisis en el contexto de sistemas de videovigilancia. Fuente: Blog de Ramón Segarra Clara. El Perfil M de ONVIF. URL: <https://ramonsegarraclara.com/el-perfil-m-de-onvif> (03/09/2024)

- Capacitar al personal de seguridad en la interpretación y uso efectivo de los datos generados³².

CASOS DE ESTUDIO Y APLICACIONES PRÁCTICAS

Smart Cities

En ciudades inteligentes, los sistemas CCTV con Inteligencia Artificial se utilizan para:

- Identificar matrículas entrantes y salientes del término municipal. Con dicha información puede saberse si un vehículo ha sido robado o usado en algún delito, si está al corriente del pago del seguro o si ha superado con éxito la Inspección Técnica de Vehículos (ITV), y si ésta está en vigor.
- Detectar y prevenir actividades delictivas en tiempo real (peleas, incendios, giros indebidos, aparcamiento en zonas prohibidas) y realizar búsquedas de personas que hayan delinquido (recuadrando la imagen del delincuente).
- Gestionar el flujo de tráfico y peatones (controlar aforos, detectar tumultos o aglomeraciones, colas de vehículo o personas, ...)
- Alertar a los ciudadanos de posibles conductas que entrañen riesgo para ellos u otros (atravesar un paso de peatones hablando por el móvil, depositar objetos en las vías que impidan el paso, ...)
- Responder rápidamente a emergencias o incidentes (caídas, peleas, incendios, ...)³³.

Comercio (retail) y Análisis de Clientes

En el sector del comercio (retail), estas tecnologías se aplican para:

- Detección de zonas de mayor o menor afluencia (mapas de calor).
- Detección de colas y tiempos de espera, con el fin de tomar medidas reactivas.

32 Fuente: <https://www.youtube.com/watch?v=n9N3B4x61AM> (05/09/2024)

33 Fuente: <https://www.altaico.es/camaras-de-vigilancia-con-inteligencia-artificial/> (05/09/2024)

- Detectar la pérdida desconocida (robo y consumo dentro del propio comercio).
- Responder rápidamente a emergencias o incidentes (caídas, peleas, incendios, ...)³⁴.
- Controlar que el personal sigue las normas en lo referente a: uso de uniforme, acceso a zonas restringidas, ausencias del puesto de trabajo, quedarse dormido, uso del móvil, prohibiciones de fumar.
- Analizar patrones de compra y comportamiento del cliente.
- Disuadir y prevenir robos y fraudes³⁵.

Seguridad Industrial

En entornos industriales, el Deep Learning en CCTV ayuda a:

- Monitorear procesos de producción para detectar anomalías.
- Asegurar el cumplimiento de protocolos de seguridad.
- Responder rápidamente a emergencias o incidentes (caídas, peleas, incendios, ...)³⁶.
- Controlar que el personal sigue las normas en lo referente a: uso de uniforme, acceso a zonas restringidas, ausencias del puesto de trabajo, quedarse dormido, uso del móvil, prohibiciones de fumar.
- Prevenir accidentes laborales³⁷.

Gestión de Infraestructuras Críticas

En instalaciones críticas como centrales eléctricas o aeropuertos, estos sistemas son cruciales para:

34 Fuente: <https://docs.euroma.es/descargas/manuales/VAIDIO-ANALITICA-VIDEO.pdf> (05/09/2024)

35 Fuente: <https://www.visiotechsecurity.com/es/noticias/365-camaras-de-seguridad-con-inteligencia-artificial> (05/09/2024)

36 Fuente: <https://www.altaico.es/camaras-de-vigilancia-con-inteligencia-artificial/> (05/09/2024)

37 Fuente: https://revistainnovacion.com/nota/10337/como_deep_learning_beneficia_el_sector_de_la_seguridad/ (05/09/2024)

- Detectar intrusiones o actividades sospechosas.
- Detectar pérdidas de objetos (objeto dejado) y sustracciones de objetos (objeto robado).
- Monitorear áreas de acceso restringido.
- Responder rápidamente a emergencias o incidentes (caídas, peleas, incendios, ...) ^{38 39}.
- Controlar que el personal sigue las normas en lo referente a: uso de uniforme, ausencias del puesto de trabajo, quedarse dormido, uso del móvil, prohibiciones de fumar.
- Prevenir accidentes laborales o industriales.

TENDENCIAS FUTURAS Y DESARROLLOS EMERGENTES

Integración con IoT y 5G

La convergencia del Deep Learning en CCTV con el Internet de las Cosas (IoT) y las redes 5G promete:

- Ubicuidad, estos sistemas podrán utilizarse en entornos remotos sin electricidad (alimentados por placas solares y batería) y conectados mediante 5G.
- Mayor conectividad y velocidad de transmisión de datos.
- Integración más fluida con otros sistemas inteligentes, dispositivos IoT⁴⁰.

Mejoras en Algoritmos de IA

Los avances continuos en algoritmos de IA están llevando a:

- Mayor precisión en la detección y clasificación de objetos.
- Más información extraída por objeto.
- Implementación de modelos orientados a la detección de comportamientos.

38 Fuente: <https://www.youtube.com/watch?v=n9N3B4x61AM> (05/09/2024)

39 Fuente: <https://docs.euroma.es/descargas/manuales/VAIDIO-ANALITICA-VIDEO.pdf> (05/09/2024)

40 Fuente: <https://www.altaico.es/camaras-de-vigilancia-con-inteligencia-artificial/> (05/09/2024)

- Capacidad para manejar escenarios más complejos y dinámicos.
- Reducción en los requisitos de potencia de procesamiento⁴¹.

Análisis Predictivo

El futuro de la inteligencia artificial en el CCTV apunta hacia capacidades predictivas más avanzadas:

- Anticipación de patrones de comportamiento potencialmente peligrosos.
- Predicción de eventos basados en análisis históricos y en tiempo real.
- Mejora en la prevención proactiva de incidentes⁴².

Ética y Regulación

A medida que estas tecnologías se vuelven más prevalentes, es probable que veamos:

- Mayor escrutinio y regulación sobre el uso de IA en vigilancia.
- Desarrollo de estándares éticos para el uso de tecnologías de reconocimiento facial.
- Aumento en las medidas de protección de la privacidad y los datos personales⁴³.

CONCLUSIÓN

La integración del Deep Learning en sistemas CCTV representa un salto cualitativo en las capacidades de videovigilancia y análisis de seguridad. Esta tecnología no solo mejora significativamente la precisión y eficiencia de los sistemas de seguridad, sino que también abre nuevas posibilidades para la

41 Fuente: <https://www.visiotechsecurity.com/es/noticias/365-camaras-de-seguridad-con-inteligencia-artificial> (05/09/2024)

42 Fuente: https://revistainnovacion.com/nota/10337/como_deep_learning_beneficia_el_sector_de_la_seguridad/ (05/09/2024)

43 Fuente: <https://www.youtube.com/watch?v=n9N3B4x61AM> (05/09/2024)

gestión inteligente de espacios y la optimización de operaciones en diversos sectores: retail, industria, gestión municipal,...

Sin embargo, con este gran poder viene una gran responsabilidad. La implementación de estas tecnologías debe realizarse con una cuidadosa consideración de las implicaciones éticas y de privacidad, asegurando un equilibrio entre la seguridad y los derechos individuales.

A medida que las capacidades de la inteligencia artificial embebida en sistemas de CCTV continúen evolucionando, podemos esperar ver sistemas de CCTV aún más sofisticados y capaces, que no solo reaccionarán a eventos, sino que los anticiparán y prevendrán de manera proactiva. El futuro de la seguridad y la vigilancia está indisolublemente ligado al avance de la inteligencia artificial, prometiendo un mundo más seguro y eficiente, siempre que se implemente y utilice de manera responsable y ética.

BIBLIOGRAFÍA

1. Smith, J. (2023). "Deep Learning in Modern CCTV Systems: A Comprehensive Analysis". *Journal of Security Technology*, 45(3), 278-295.
2. Chen, L., & Wang, H. (2022). "Advancements in AI-Powered Video Surveillance". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8), 1567-1583.
3. Rodriguez, A., et al. (2024). "Neural Networks for Real-Time Video Analysis in Security Applications". *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, 89-103.
4. Thompson, E. (2023). "Ethical Considerations in AI-Enhanced Surveillance Systems". *AI & Society*, 38(2), 345-360.
5. Liu, Y., & Zhang, X. (2022). "Deep Learning Algorithms for Object Detection in CCTV Footage". *Computer Vision and Image Understanding*, 215, 103329. 

PROCESAMIENTO DE LENGUAJE NATURAL (PLN)

💬 Joaquín Rieta y Juan Martínez

1. INTRODUCCIÓN

La interacción entre el lenguaje humano y las máquinas es compleja pues implica no solo comprender el significado de las palabras, sino también el contexto, la gramática y las emociones detrás del texto. Imaginemos la ironía y las cuestiones siempre presentes dentro del ámbito sociocultural que hace muy difícil inclusión a la hora de comunicarnos entre humanos de diferentes culturas o regiones. Pues bien, para dar solución a este reto nace el Procesamiento de Lenguaje Natural (PLN, por sus siglas en inglés) que es dentro del campo de la inteligencia artificial tiene por objetivo dar solución a esa interacción entre las máquinas y el lenguaje humano.

2. ¿QUÉ ES EL PROCESAMIENTO DE LENGUAJE NATURAL?

El procesamiento de lenguaje natural se centra en hacer que las máquinas nos entiendan y podernos comunicar de manera efectiva con ellas, es decir, que entiendan el lenguaje humano en su forma más amplia y compleja. Esto incluye analizar el texto para extraer su significado real, responder a las preguntas, traducción de idiomas, y mucho más. Basándose en el uso de técnicas matemáticas y modelos de aprendizaje automático, se consigue que las máquinas procesen grandes cantidades de datos y obtengan información relevante e inteligible por un humano.

Por ejemplo, cuando le decimos a nuestro asistente virtual “¿Qué tiempo va a hacer hoy?”, entra en acción el PLN que hace que el sistema interprete la pregunta y te dé una respuesta relevante, como “Será soleado con una temperatura máxima de 25°C y mínima de 15°C”.

3. COMPONENTES CLAVE Y TÉCNICAS

El procesamiento de lenguaje natural de las máquinas funciona a través de una programación basada en varias tareas y técnicas que le permiten

descomponer el lenguaje humano en partes más manejables para su procesamiento. Vamos a citar esos componentes clave:

- **Tokenización:** Se trata de dividir el texto en palabras o frases individuales (llamados "tokens").
- **Análisis Sintáctico:** Que determina la estructura gramatical de las oraciones.
- **Análisis Semántico:** Consiste en entender el significado de las palabras dentro del contexto de una oración.
- **Reconocimiento de Entidades:** Permite identificar nombres de personas, lugares, organizaciones, fechas, etc.
- **Resolución de Ambigüedades:** Resuelve el significado cuando una palabra puede tener múltiples interpretaciones.
- **Análisis de Sentimiento:** Determina si un texto expresa emociones positivas, negativas o neutras.

Estos componentes clave forman parte del procesamiento del lenguaje que junto a las ya citadas técnicas matemáticas y modelos de aprendizaje automático consiguen que nos podamos comunicar de una manera efectiva y útil con las máquinas. Entre las técnicas para desarrollar los PLN podemos encontrarnos con simples enfoques basados en reglas hasta métodos más avanzados que utilizan aprendizaje automático. Para ilustrar un poco en qué consiste citamos algunas:

- **Bag of Words (Bolsa de Palabras):** Este método simplifica el texto al tratar las palabras como elementos independientes, ignorando el orden y el contexto. Es como una sopa de letras inconexa.
- **TF-IDF (Frecuencia del Término - Frecuencia Inversa del Documento):** Una técnica para medir cuán importante es una palabra en un texto o documento específico. Por ejemplo, los verbos frente a los artículos.
- **Modelos de Word Embeddings:** Métodos como Word2Vec o GloVe crean representaciones vectoriales de palabras que capturan el significado contextual.

- **Modelos Basados en Transformadores:** Los modelos avanzados, como BERT o GPT, utilizan transformadores para entender el contexto de las palabras en el texto.

Esos componentes junto con las técnicas hacen “la magia” para que éstas interactúen con nosotros de forma útil y legible.

4. APLICACIONES

Son muchos los campos de aplicación en los que el PLN aparece, vivimos en un mundo conectado con máquinas, casi sin darnos cuenta el PLN está en cada interacción humano – máquina de nuestro día a día. Entre las aplicaciones más comunes figuran:

- **Traducción Automática:** Servicios como Google Translate lo utilizan para traducir texto entre diferentes idiomas.
- **Asistentes Virtuales:** Siri, Alexa, y Google Assistant lo usan para comprender comandos de voz y responder a preguntas.
- **Análisis de Sentimiento:** Las empresas lo utilizan para analizar opiniones en redes sociales o reseñas de productos y determinar si los comentarios son positivos o negativos.
- **Resúmenes Automáticos:** Herramientas que pueden resumir artículos largos en unas pocas frases clave.
- **Chatbots y Avatares:** Muy generalizados en sitios web que, mediante PLN, pueden interactuar con los usuarios y proporcionar asistencia en tiempo real.

En nuestra rutina diaria, por ejemplo, lo encontramos en los servicios de atención al cliente, muchas empresas están utilizando Chatbots o Avatares para atender a los clientes de manera automática, interpretando las preguntas de los usuarios y ofreciendo soluciones rápidas sin la intervención de un humano. (Banca, Suministros, Servicios, Administración pública, ...) esto no sería posible sin el procesamiento de lenguaje natural.

Una empresa de marketing en el trasfondo, a través de aplicaciones, usa PLN para analizar miles de comentarios en Twitter y detectar tendencias

sobre la opinión pública respecto a un producto recientemente lanzado. Sin el PLN sería imposible por tiempo y coste.

5. LIMITACIONES Y DESAFÍOS

Quién no se ha encontrado llamando por teléfono y oyendo una voz que te demanda pulsar tal o cual tecla y ya últimamente que digas tu consulta. Todo es procesamiento de lenguaje natural que ha facilitado la mejora de servicios y algún que otro enfado por parte del usuario por citar limitaciones que también las tiene.

Se ha avanzado mucho en los últimos años, pero aún enfrenta varios desafíos:

- **Ambigüedad:** El lenguaje humano es inherentemente ambiguo, lo que significa que una misma palabra puede tener varios significados dependiendo del contexto y del idioma. Esto sigue siendo un reto para los modelos de PLN.
- **Lenguaje Figurativo:** Las máquinas tienen dificultades para entender el sarcasmo, la ironía, las metáforas o los juegos de palabras.
- **Multilingüismo:** Aunque se ha mejorado la traducción automática, sigue siendo complicado lograr una comprensión perfecta de todos los idiomas y dialectos.
- **Ruido en los Datos:** Los datos de los que aprenden los modelos de PLN a menudo contienen errores o sesgos que pueden afectar al rendimiento del modelo.

Todo esto es un desafío para los PLN que sin duda requerirá de tiempo pero la tendencia como veremos más adelante es a salvar esos retos.

6. DESAFÍOS ÉTICOS EN EL PROCESAMIENTO DE LENGUAJE NATURAL

Tenemos que hacer mención que el uso del procesamiento del lenguaje natural en nuestra sociedad está planteando debates sobre desafíos éticos que deben ser abordados sin lugar a dudas, por citar a modo de resumen tenemos los referentes a la:

- **Privacidad de los Datos:** Muchas aplicaciones requieren de grandes cantidades de datos, que a menudo incluyen información personal, esto plantea las lógicas preocupaciones sobre la privacidad, su manejo y el uso adecuado de los mismos.
- **Sesgos en los Modelos:** Los modelos aprenden los sesgos presentes en los datos con los que son entrenados, lo que puede llevar a la generación de contenido discriminatorio o prejuicios, ya arraigados en nuestras sociedades a lo largo del tiempo.
- **Manipulación de Información:** Toda esa inmensa masa de datos es muy valiosa y puede ser utilizada para generar noticias falsas o contenido engañoso, lo que puede afectar a la percepción pública y la toma de decisiones en favor de una tendencia que beneficie a un tercero interesado.

7. HERRAMIENTAS Y PLATAFORMAS

Para trabajar con NPL existen muchas herramientas y plataformas disponibles que están constantemente apareciendo. A modo somero citamos algunas de las más populares como son:

- **SpaCy:** Una biblioteca en Python para el procesamiento eficiente de texto en múltiples idiomas.
- **NLTK (Natural Language Toolkit):** Un conjunto de bibliotecas y programas en Python para trabajar con texto en lenguaje natural.
- **Hugging Face Transformers:** Una plataforma que proporciona modelos avanzados de PLN preentrenados, como BERT y GPT.
- **Google Cloud Natural Language API:** Un servicio en la nube que permite a los desarrolladores analizar texto, extraer entidades y realizar análisis de sentimientos.

8. FUTURO DE LOS PROCESAMIENTO DE LENGUAJE NATURAL

Los avances se producen cada vez en un tiempo más corto siempre con el foco en mejorar aún más las capacidades de las máquinas para entender y generar lenguaje humano, prácticamente cada año se producen avances, vamos a citar tres tendencias hacia las que evoluciona el PLN:

- **Modelos Multimodales:** Desarrollos que combinan texto, imágenes y sonidos para permitir que las máquinas entiendan el contexto de manera más integral.
- **Mejor Compresión de Contexto:** Intentan captar mejor el significado más profundo y las emociones detrás del texto, lo que permitirá una interacción más natural con los usuarios.
- **Mayor Adaptabilidad:** Los sistemas de PLN serán más personalizables y adaptables a las necesidades específicas de cada usuario o aplicación. Se está hablando de personalización por tipos, profesiones y sectores.

El Procesamiento de Lenguaje Natural ha revolucionado la forma en que interactuamos con las máquinas, nos permite que éstas nos entiendan y respondan a nuestro lenguaje con una precisión y una naturalidad cada vez mayores. Desde los asistentes virtuales hasta el análisis de datos en tiempo real, tiene aplicaciones que abarcan una amplia gama de industrias. Sin embargo, es crucial abordar los desafíos éticos y técnicos para garantizar que esta tecnología sea utilizada de manera justa y beneficiosa para la sociedad.

A medida que la tecnología avanza, veremos más innovaciones que harán que las interacciones con la IA sean aún más naturales y eficaces. 📄

GRANDES MODELOS DE LENGUAJE (LLM)

💬 Joaquín Rieta y Juan Martínez

1. INTRODUCCIÓN

Una de las claves por las cuales se ha popularizado el interés por la inteligencia artificial dentro del público general en tan sólo dos años ha sido gracias a la evolución y el desarrollo de los llamados “Grandes Modelos de Lenguaje” (LLM, por sus siglas en inglés).

Los LLM son una tecnología emergente en el campo de la inteligencia artificial (IA) que ha transformado la forma en que interactuamos con el lenguaje de una manera sencilla y práctica permitiendo en muchos casos el acceso al uso más general de la inteligencia artificial en diversos campos.

Estos modelos son capaces de responder preguntas, escribir párrafos e historias, generar código informático, y mucho más. Se entrenan con grandes volúmenes de datos (millones o incluso miles de millones de textos) para predecir cuál será la siguiente palabra en una oración, generando así texto coherente, lo que les permite entender, generar y manipular texto de manera avanzada.

Para ayudar a entender qué son, cómo funcionan, y qué tipos de aplicaciones tienen en el mundo real vamos a desgranar brevemente que son, cuál es su arquitectura, sus capacidades, sus limitaciones, sus desafíos, su uso responsable y ético, para acabar con una visión de hacia donde se encaminan en un futuro no muy lejano.

2. ¿QUÉ ES UN GRAN MODELO DE LENGUAJE?

Los grandes modelos de lenguaje están basados en una arquitectura conocida como “transformador” (Transformers) que les permite aprender relaciones complejas entre palabras y conceptos. Existen actualmente varios, los más conocidos como GPT-4 de la empresa OpenAI o GEMINI de Google entre otros.

Los Transformers son un tipo de redes neuronales que se diferencian de otros tipos gracias a que pueden procesar grandes cantidades de datos a la vez, en vez de hacerlo de manera secuencial. Esto los hace mucho más

rápidos y eficientes para tareas que implican analizar y generar texto, como por ejemplo, la traducción de idiomas o la escritura automática.

¿Cómo funcionan los Transformers?

El núcleo de los Transformers es algo llamado "mecanismo de atención". Esto permite que el modelo decida qué partes de la información son más importantes a medida que procesa una secuencia de datos. Por ejemplo, si el modelo está traduciendo una frase del español al inglés, prestará atención a las palabras clave que influyen más en la traducción correcta.

A diferencia de las redes recurrentes, que procesan una palabra o un dato a la vez, los Transformers procesan todos los datos de la secuencia al mismo tiempo. Esto hace que sean mucho más rápidos para entrenar y usar.

Si el LLM recibe el inicio de una oración como "El niño está...", es capaz de continuarla con "...durmiendo y tranquilo" basándose en patrones que ha aprendido de su entrenamiento.

3. ARQUITECTURA Y ENTRENAMIENTO

Los LLM están contruidos bajo la arquitectura de los Transformers, desarrollada en su día por Google (2007), sobre redes neuronales profundas. Esto permite que los modelos manejen texto de manera más eficiente y entiendan el contexto a largo plazo, algo crucial para tener coherencia y precisión en el lenguaje, cuestiones muy importantes.

Los modelos más grandes, como GPT-4, por ejemplo, contienen miles de millones de parámetros ajustables que permiten mejorar el rendimiento y la precisión en tareas complejas.

Para su entrenamiento se utilizan millones de libros, artículos, páginas web, y otros recursos textuales. El proceso de entrenamiento requiere disponer de unos centros con gran capacidad de computación por lo que suelen ser bastante costosos pudiendo durar en el tiempo semanas o meses.

4. CAPACIDADES DE LOS GRANDES MODELOS DE LENGUAJE

Gracias al entrenamiento masivo, los LLM son capaces de realizar una infinidad de tareas con el lenguaje, por ejemplo, entre otras para:

- **Generar Texto:** Pueden escribir textos originales a partir de una simple solicitud.
- **Traducción Automática:** Pueden traducir de un idioma a otro con precisión.
- **Resumen de Textos:** Resumen artículos largos en unos pocos párrafos.
- **Respuesta a Preguntas sobre una temática:** Son capaces de responder preguntas basándose en su conocimiento almacenado hasta una determinada fecha.
- Como **Asistentes Virtuales** son usados habitualmente en aplicaciones, tan comunes hoy en día, como son Siri o Alexa, los LLM permiten entender el lenguaje natural dentro de un contexto cotidiano realizando tareas.

Los LLM ya estaban conviviendo en nuestro entorno casi sin darnos cuenta en muchos aspectos de nuestra vida cotidiana. Son utilizados en:

- **Asistentes Virtuales y Chatbots:** Servicios como ChatGPT pueden responder preguntas, ayudar con tareas o realizar recomendaciones personalizadas.
- **Motores de Búsqueda:** Google utiliza LLM para mejorar la comprensión de las consultas de los usuarios, ofreciendo resultados más precisos.
- **Generación de Contenido:** Herramientas como Jasper.ai usan LLM para generar artículos, blogs o incluso escribir código.
- **Educación:** Personalizando y adaptando el aprendizaje a un nivel determinado de conocimiento requerido, respondiendo preguntas y explicando conceptos difíciles a demanda.

En definitiva, los campos de uso son ilimitados pudiéndose aplicar a cualquier tipo de sector e industria. Dentro del área del Marketing Digital, podemos crear contenido automatizado, generar eslóganes publicitarios o recomendaciones de productos, diseñar marcas y actuar en las redes sociales optimizándolas (SEO).

En atención al cliente, quien no se ha encontrado con los Chatbots que interactúan de manera fluida con los clientes, resolviendo problemas o guiando al usuario en tareas (telefonía, banca, administración pública, ...). Por ejemplo:

Una aerolínea puede utilizar un Chatbots basado en LLM para ayudar a los usuarios a cambiar fechas de vuelos, todo mediante lenguaje natural.

En el área de la Medicina se utilizan los LLM para análisis de datos médicos, generación de informes o incluso asistencia en diagnósticos. Por ejemplo: Un hospital o clínica puede utiliza un LLM para ayudar a los médicos en investigaciones y la toma de decisiones clínicas. La investigación científica es otra área en la que se utilizan.

5. LIMITACIONES

A pesar de su poder, evidente, los LLM tienen ciertas limitaciones que se deben tener muy presentes, destacamos tres aspectos:

- **Sesgos en los Datos:** Dado que los modelos aprenden de grandes cantidades de datos textuales de internet, éstos pueden aprender sesgos raciales, de género o culturales que se encuentran en esos datos.
- **Privacidad:** Algunos modelos pueden almacenar información sensible o privada, lo que plantea riesgos de seguridad que se deben tener muy en cuenta en el uso.
- **Generación de Información Incorrecta:** Aunque son buenos para generar texto coherente, a veces los modelos pueden dar respuestas incorrectas o hacer afirmaciones falsas, ya que no tienen una “comprensión” del mundo, sino que se basan en correlaciones de palabras.

Tener en cuenta estos aspectos limitantes es importante por lo que se hace necesaria una adecuada formación que nos va a permitir aprovechar de manera más útil y consciente estos modelos.

6. DESAFÍOS ÉTICOS EN EL USO DE LLM

La creación y uso de LLM conlleva una serie de desafíos éticos. Como todo en la vida tiene sus dos caras, una positiva y otra negativa. La desinformación puede ser capaz de generar contenido convincente, que oriente hacia un lado sesgado e interesado por un tercero malicioso, lo que podría ser utilizado para crear noticias falsas o engaños.

Tenemos ya al orden del día “Deepfakes”, estos modelos pueden ser utilizados para generar texto, imágenes o incluso videos que imiten a personas reales, lo que plantea problemas serios relacionados con la manipulación de la información o sustitución de personalidad.

Otro desafío al que nos llevan es el relativo a los recursos humanos siendo obvio el impacto en el empleo que supone la automatización de ciertas tareas, como la escritura de contenido o la atención al cliente que podría reducir la demanda de trabajo humano en estas áreas que afectarán más a algunos puestos y sectores que a otros pero que sin duda impactará en todo el abanico social creando nuevos puestos y dando por eliminados totalmente otros.

Es fundamental establecer marcos regulatorios y principios éticos para guiar el desarrollo y uso de los LLM de manera responsable para esos desafíos, así como anticipar su impacto social en el área laboral que será importante.

7. ¿CÓMO USAR UN LLM DE FORMA RESPONSABLE?

Para garantizar el uso ético y útil de los LLM, es importante tener en cuenta algunas recomendaciones como la supervisión humana, aunque los LLM son potentes, es necesario que el contenido generado sea revisado por humanos, especialmente en áreas sensibles como la medicina o la ley. El filtrado de contenidos inapropiados se hace necesario implementan consensos de mecanismos que detecten y filtren respuestas que promuevan la violencia, la discriminación o la información falsa.

Se requiere transparencia e informar a los usuarios cuando estén interactuando con un LLM y explicar cómo se han entrenado estos modelos, cuestión que las compañías deben de tener presentes.

8. FUTURO DE LOS GRANDES MODELOS DE LENGUAJE: LAM (LARGE ACTION MODELS).

La investigación de los LLM continúa mejorando exponencialmente su capacidad, precisión y eficiencia. Se espera que estos modelos se vuelvan más multimodales, es decir, que puedan manejar no solo texto, sino también imágenes, video y sonido, lo que les permitirá realizar tareas aún más complejas y diversas. Dentro de los desarrollos esperados pasaran por una mejor

comprensión contextual, mayor personalización y ampliación a más campos desde la creatividad hasta la toma de decisiones.

El futuro de los Grandes Modelos de Lenguaje (LLM) sigue siendo prometedor, con avances constantes que permiten que estos sistemas sean cada vez más precisos, eficientes y aplicables en una gama aún más amplia de industrias. Un área que está ganando importancia son los grandes modelos de acción (LAM por su siglas en inglés), que van más allá del procesamiento de texto para incluir otros tipos de datos y ampliar las capacidades de la inteligencia artificial.

Un Large Action Model (LAM) es un modelo de inteligencia artificial que puede entender y ejecutar tareas complejas al traducir las intenciones humanas en acciones. Son modelos que permiten el análisis de grandes volúmenes de información, ayudando a resolver problemas complejos en sectores como la salud, la industria o las finanzas. Su entrenamiento se basa en redes neuronales profundas que permiten manejar diferentes tipos de datos, como imágenes y señales temporales. Gracias a este entrenamiento, los LAM son capaces de realizar tareas como la predicción, clasificación de datos y reconocimiento de patrones, ampliando el abanico de aplicaciones posibles.

Sin embargo, los LAM también presentan desafíos, como la necesidad de grandes cantidades de datos para ser efectivos y su dependencia de infraestructura tecnológica costosa y compleja. Además, su capacidad para generalizar patrones puede verse afectada por sesgos en los datos de entrenamiento, lo que plantea cuestiones éticas importantes. A pesar de estas limitaciones, su uso responsable —con supervisión humana y marcos regulatorios adecuados— permitirá aprovechar el potencial de los LAM en combinación con los LLM.

Como resultado, se espera que el futuro de los LLM esté cada vez más vinculado a la integración de modelos multimodales, que combinen texto, imágenes, sonido y otros tipos de datos en un solo sistema capaz de manejar tareas más diversas y complejas. Esto abrirá nuevas oportunidades para la automatización de procesos en industrias que van desde la educación hasta la investigación científica y el entretenimiento.

Como conclusión podemos decir que los Grandes Modelos de Lenguaje son una de las herramientas más poderosas en el campo de la inteligencia

artificial que han venido para quedarse y que cambiarán nuestra manera de interactuar y de trabajar. Su capacidad para procesar y generar lenguaje natural tiene un impacto significativo y transversal como la atención al cliente, la educación, la medicina y el marketing.

Sin embargo, es esencial abordar los desafíos éticos y las limitaciones técnicas para maximizar sus beneficios y minimizar los riesgos.

A medida que estos modelos evolucionan en el tiempo, es fundamental que tanto los desarrolladores como los usuarios adopten un enfoque cada vez más responsable y ético en su uso, para que su impacto en la sociedad sea lo más positivo y equitativo posible.

Para resumir vemos una tabla resumen de los algunos aspectos clave sobre los LLM y vocabulario de los LAM.

ASPECTO	DESCRIPCIÓN
Definición	Modelos de IA entrenados con grandes cantidades de datos textuales para procesar y generar lenguaje natural.
Ejemplos de Modelos	GPT-4 (OpenAI), GEMINI (Google), LLaMA (Meta), ...
Tareas Comunes	Traducción automática, generación de texto, respuesta a preguntas, resumen de textos, etc.
Arquitectura	Basados en redes neuronales profundas, especialmente en arquitecturas de transformadores (Transformers).
Entrenamiento	Requiere grandes volúmenes de datos textuales y poder computacional significativo.
Capacidades	Entender el contexto, generar respuestas coherentes, realizar inferencias, manipular lenguaje de manera compleja.
Aplicaciones	Asistentes virtuales, motores de búsqueda, Chatbots, análisis de sentimiento, creación de contenido automatizado, etc.
Limitaciones	Sesgos inherentes a los datos de entrenamiento, alta demanda de recursos computacionales, generación de contenido erróneo o sesgado.

Desafíos Éticos	Propagación de desinformación, sesgo de género o racial, privacidad de datos, uso en la creación de Deepfakes.
Ejemplos de Uso	ChatGPT para asistencia en tareas cotidianas, BERT para mejorar la precisión en la búsqueda de Google.
Evolución	Desde modelos más simples y específicos hasta modelos multitarea y multimodales con capacidades avanzadas.
Tamaño de Modelo	Varía desde cientos de millones hasta cientos de miles de millones de parámetros, dependiendo del modelo.

VOCABULARIO – LAM

- **Lenguajes de Aprendizaje Automático (LAM)**
Large Action Model (LAM): Un LAM es un modelo de IA entrenado con una gran cantidad de datos de acciones e instrucciones humanas, que puede comprender y responder a comandos en lenguaje natural y ejecutar tareas complejas de múltiples pasos, interactuando con herramientas y aplicaciones de software.
- **Red Neuronal Convolutiva (CNN)**
Convolutional Neural Network (CNN): Arquitectura de red neuronal diseñada para procesar datos con una estructura de cuadrícula, como imágenes, y detectar patrones.
- **Red Neuronal Recurrente (RNN)**
Recurrent Neural Network (RNN): Tipo de red neuronal diseñada para trabajar con datos secuenciales, como series temporales o lenguaje, manteniendo información de estados anteriores.
- **Sesgos en los Datos**
Bias in Data: Tendencias o prejuicios en los datos de entrenamiento que pueden afectar el rendimiento y la imparcialidad de los modelos.
- **Generalización de Patrones**
Pattern Generalization: Capacidad de un modelo para aplicar lo aprendido de datos de entrenamiento a nuevos datos no vistos previamente.
- **Descentramiento de Gradiente**

Gradient Descent: Algoritmo de optimización utilizado en el entrenamiento de redes neuronales para ajustar los parámetros del modelo y minimizar el error.

- **Parámetros Ajustables**

Adjustable Parameters: Variables en un modelo que se ajustan durante el entrenamiento para optimizar su rendimiento.

- **Detección de Anomalías**

Anomaly Detection: Proceso mediante el cual los modelos detectan irregularidades o desviaciones de patrones esperados en los datos.

- **Filtrado de Sesgos**

Bias Filtering: Técnicas para identificar y mitigar sesgos en los datos de entrenamiento, mejorando la equidad de los modelos.

- **Supervisión Humana**

Human Oversight: Intervención de expertos humanos para revisar y validar las decisiones o resultados generados por los modelos de aprendizaje automático.

- **Transformadores**

Transformers: Tipo de arquitectura de red neuronal que procesa secuencias de datos de manera más eficiente, comúnmente utilizada en grandes modelos de lenguaje (LLM).

- **Multimodalidad**

Multimodality: Capacidad de un modelo para procesar diferentes tipos de datos, como texto, imágenes y sonido, de forma simultánea.

- **Generación de Contenido**

Content Generation: Habilidad de los modelos para crear texto, imágenes u otros tipos de contenido basados en una entrada dada.

- **Computación Distribuida**

Distributed Computing: Método de utilizar múltiples sistemas informáticos en paralelo para procesar grandes volúmenes de datos durante el entrenamiento de modelos.

- **Transparencia en Modelos**

Model Transparency: Práctica de informar a los usuarios sobre cómo se entrenan los modelos, qué datos se utilizan y cómo se toman decisiones.

- **Protección de la Privacidad**

Privacy Protection: Conjunto de prácticas para asegurar que los datos personales utilizados en el entrenamiento de modelos se mantengan seguros y privados. 📄

VISIÓN POR COMPUTADOR (CV: COMPUTER VISION)

💬 Ivana Aguilar

INTRODUCCIÓN

La visión por computador es un campo de la inteligencia artificial que permite a las máquinas interpretar y comprender imágenes o videos. Se basa en técnicas de procesamiento de imágenes, como la extracción de características (bordes, texturas, formas), segmentación y reconocimiento de patrones.



Los algoritmos, como redes neuronales convolucionales (CNN), juegan un papel crucial en tareas como la clasificación de imágenes, la detección de objetos y el seguimiento. El aprendizaje profundo ha mejorado notablemente la precisión de estas tareas, permitiendo a los sistemas aprender de grandes volúmenes de datos etiquetados. Además, la visión por computador utiliza modelos matemáticos y estadísticos para reconstruir y analizar escenas tridimensionales, logrando que las máquinas entiendan el entorno visual de forma similar a los humanos.

APRENDIZAJE PROFUNDO Y VISIÓN POR COMPUTADOR

Como ya se mencionó en el capítulo 2 de esta guía, 'Tipos de IA', Yan LeCun fue el pionero en el desarrollo de las redes neuronales convolucionales (*Convolutional Neural Networks, CNN*). En los años 80, las redes neuronales eran

vistas por muchos investigadores con mucho escepticismo por 3 motivos: eran difíciles de entrenar, se quedaban atrapadas en mínimos locales y los recursos computacionales eran limitados. No obstante, la fascinación de LeCun por las redes neuronales fue lo que lo hizo perseverar en su investigación, consiguiendo afianzarlas a finales de los 80 y a principios de los 90, marcando un hito en el desarrollo de la inteligencia artificial.

Las redes neuronales convolucionales (RNC), son una clase de redes neuronales diseñadas específicamente para procesar datos estructurados, en donde los elementos se distribuyen en un formato matricial o de grilla, como las imágenes. Este tipo de redes imita el funcionamiento del cerebro humano en la forma en que procesa la información visual.

Dos propiedades interesantes que tienen las RNC son:

- Los patrones que estas redes aprenden son invariantes en translación, esto quiere decir, que después de aprender un determinado patrón, son capaces de reconocer este patrón en cualquier parte de la imagen¹.
- Estas redes aprenden jerarquías espaciales de los patrones; esto es, que en una primera capa de convolución aprenderá pequeños patrones locales, mientras que una segunda capa de convolución aprenderá patrones más grandes hechos de las características de las primeras capas y así sucesivamente. Esto es lo que las hace muy eficientes ya que aprenden el complejo y abstracto concepto visual².

Los componentes básicos de una red neuronal convolucional son dos capas que pueden expresarse como grupos de neuronas especializadas en dos operaciones: convolución y agrupamiento (*pooling*). La capa convolucional aplica simultáneamente múltiples filtros entrenables a sus entradas, lo que la hace capaz de detectar múltiples características en cualquier parte de sus

1 Libro: *'Deep learning con Python'*. Ediciones Anaya multimedia (grupo ANAYA S.A) 2020. Chollet F.

2 Libro: *'Aprende Machine Learning con Scikit-Learn, Keras y TensorFlow'*. O'Reilly 2009. Aurélien G.

entradas, mientras que el objetivo de la capa de *pooling* es sub-samplear la imagen de entrada para de esta forma reducir la carga computacional, el uso de la memoria y el número de parámetros. De este modo evita el riesgo de sobre entrenamiento de la red.

Además de estas capas (*layers*) que forman la estructura principal de una RNC, se hacen uso de otras capas: densa, *max pooling2D*, *flatten* y *dropout*.

Debido al coste computacional que suponía el entrenar una RNC, en 2010 comenzó a ganar atención el concepto de transferencia de conocimiento. Un hito significativo lo realizó AlexNet (en 2012), cuando mostró la efectividad de utilizar modelos pre-entrenados en el conjunto de datos ImageNet. Desde entonces, la transferencia del conocimiento se ha consolidado como una técnica clave en el desarrollo de modelos de aprendizaje profundo.

¿QUÉ ES LA TRANSFERENCIA DEL CONOCIMIENTO O TRANSFER LEARNING?

Con la transferencia de conocimiento o *transfer learning* lo que se pretende es seleccionar un modelo de red neuronal predefinida, cuyos pesos han sido previamente entrenados con otro tipo de imágenes en tareas similares de clasificación y de la cual se han obtenido buenos resultados. Con esto se evita tener que definir su arquitectura desde cero (*from scratch*) así como también realizar un ajuste fino (*fine tuning*) para que los pesos de la red no se inicien de forma aleatoria³. En resumen, el uso de un modelo pre-entrenado simplifica el proceso de diseño y mejora la estabilidad del entrenamiento, evitando la aleatoriedad inicial que ocurre en entrenamientos desde cero.

Una de las técnicas más utilizadas en la reutilización de modelos es el ajuste fino (*fine tuning*).

El **ajuste fino** es una técnica complementaria, ampliamente utilizada en la transferencia del conocimiento. Consiste en hacer un ajuste más fino de las representaciones más abstractas que forman parte del modelo que se está reutilizando como base. Esto se consigue congelando los valores de los pesos (coeficientes de las primeras capas convolucionales), así como reentrenando

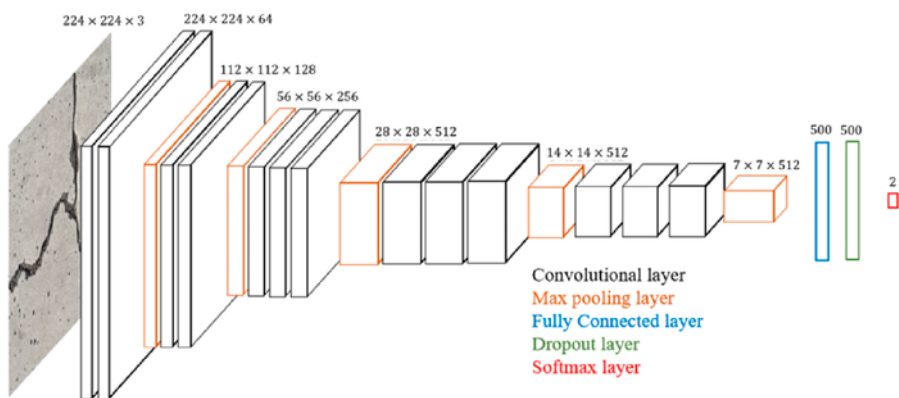
3 Diseño y desarrollo de software para el análisis de patrones microscópicos de crecimiento tumoral en pacientes con carcinoma vesical. Aguilar I.

únicamente los pesos de las últimas capas para incorporar en la arquitectura la información relativa a las imágenes específicas del problema bajo estudio.

Para entender mejor su funcionamiento, hay que tener en cuenta el nivel de generalización y la posición que tengan dichas capas (de convolución específicas) en el modelo ya que, dependiendo de esto último, interesará modificarlas en mayor medida o no. Así pues, si lo que queremos que detecte nuestra red es un objeto o un patrón determinado, se puede asumir que, las capas iniciales de la red extraerán: los bordes, texturas, colores, etc. Las capas intermedias extraerán conceptos más abstractos como pueden ser: la forma que tenga el objeto o patrón a buscar. Y las capas finales, extraerán conceptos más abstractos que nos van a permitir construir un concepto del objeto o patrón presentado.

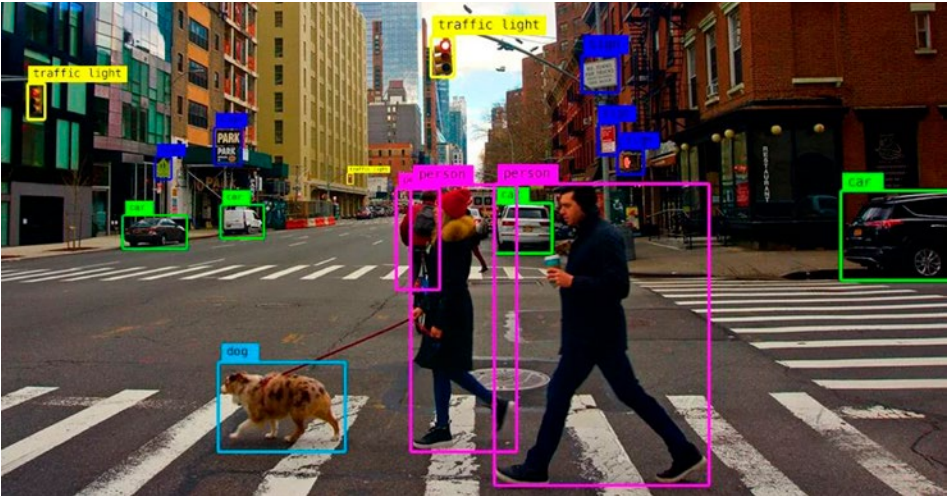
Cuando se hace una modificación (tuneo) en las capas iniciales e intermedias del modelo reutilizado, a este ajuste fino se le conoce como afinación profunda (deep tuning) y cuando la modificación es en las capas finales se le conoce como afinación superficial (shallow tuning).

Algunos de los modelos más utilizados en redes neuronales convolucionales son: LeNet, AlexNet, VGG, ResNet y Inception. Estos modelos han sido fundamentales en el avance de la visión por computador, mejorando la precisión y eficiencia en tareas como el reconocimiento de imágenes, detección de objetos y segmentación. El aporte más destacado de cada una de ellas fue: LeNet fue pionero en el reconocimiento de dígitos; AlexNet revolucionó el campo en 2012 con su éxito en ImageNet; VGG simplificó la arquitectura



de capas; ResNet introdujo conexiones residuales para entrenar redes profundas, e Inception optimizó la eficiencia computacional.

En la actualidad, una de las arquitecturas más influyentes en visión por computador es YOLO (*You Only Look Once*), diseñada para el reconocimiento y detección de objetos en tiempo real. A diferencia de los métodos tradicionales que utilizan las RNC estándar (dividir la tarea en múltiples etapas), YOLO, procesa la imagen completa, dividiéndola en una cuadrícula y prediciendo múltiples cajas delimitadoras y probabilidades de clase de manera simultánea. Esto la hace extremadamente rápida y eficiente. Desde 2016, YOLO continúa mejorando sus algoritmos de detección en tiempo real. Sus versiones más actuales YOLOv6, YOLOv7 y YOLOv8 han demostrado avances significativos en eficiencia y escalabilidad.



APLICACIONES PRÁCTICAS DE LA VISIÓN POR COMPUTADOR

El campo de la visión por computador ha conseguido que la sociedad sea más partícipe en su evolución y mejora. La velocidad a la que se comparten datos (imágenes), en las distintas plataformas, principalmente en las redes sociales, es realmente asombrosa. En la actualidad, esto permite que se desarrollen aplicaciones casi personalizadas para diversos modelos de negocios.

La siguiente tabla presenta un análisis de las tendencias actuales de las aplicaciones de visión por computador mostrando las categorías y los porcentajes de uso.

Resumen de Categorías y Porcentajes

Categoría	Porcentaje (%)
Seguridad y Vigilancia	25-30
Automatización Industrial	20-25
Sector de la Salud	15-20
Automóviles Autónomos	15-20
Retail y Comercio Electrónico	10-15
Agricultura de Precisión	5-10
Deportes y Entretenimiento	5-10
Aplicaciones Financieras y Banca	5-10
Smart Cities y Transporte Público	5-10

Tabla 1: Síntesis basada en una comprensión general de las aplicaciones en visión por computador y su relevancia en distintos sectores.

En cuanto a su impacto general, podemos apreciar lo siguiente.

En el aspecto económico, la visión por computador impulsa la productividad, mejora la eficiencia y reduce costes en diversos sectores. En el ámbito social, incrementa la seguridad, ofrece experiencias personalizadas y simplifica procesos complejos, aunque también plantea retos éticos y de privacidad. En el ámbito ambiental, contribuye a una gestión más sostenible en la agricultura y optimiza la movilidad urbana.

Entre las aplicaciones que más destacan en este campo se encuentra: el reconocimiento facial, la detección de objetos a tiempo real y la mejora de los sistemas de apoyo al diagnóstico (AID) en medicina.

El reconocimiento facial ha facilitado el análisis de microexpresiones faciales, lo cual es fundamental para estudiar las emociones dentro del lenguaje no verbal. Además, esta tecnología ha permitido la recolección de datos biométricos mediante el uso de mallas faciales, proporcionando información precisa sobre las características y movimientos del rostro para diversas aplicaciones. Estos procedimientos también se han utilizado para obtener información del iris, considerado como un dato biométrico altamente preciso y único, comparable e incluso superior a las huellas dactilares. Estudios científicos demuestran que su estructura es 10 veces más detallada que la

de una huella. Su probabilidad de coincidencia entre dos personas es de 1 en 10^{78} . Esta precisión se debe a su compleja y estable estructura, que no cambia significativamente a lo largo de la vida.

En cuanto a la detección de objetos, la red neuronal YOLO en sus distintas versiones ha ido mejorando la latencia de detección a tiempo real de objetos cotidianos. Su modelo YOLOv3 detecta hasta 80 clases y su más reciente modelo YOLOv8x detecta hasta 150 objetos (80 clases, más si es entrenado con otros conjuntos de datos), con una precisión del 55-56% y velocidad de 40-60 FPS (*frames* por segundo).

La comparativa entre estos dos modelos sería la siguiente:

Comparación: YOLOv3 vs. YOLOv8

Característica	YOLOv3	YOLOv8
Cantidad de clases (COCO dataset)	80 clases	80 clases (más si es entrenado con otros datasets)
Cantidad de instancias por imagen	Ilimitado (dependiendo de la imagen y su complejidad)	Ilimitado (mejor rendimiento que YOLOv3)
Rendimiento con objetos pequeños	Bueno, pero inferior a YOLOv8	Superior, mejor detección de objetos pequeños y cercanos

Por último, en los sistemas AID, la combinación de visión por computador y el avance en componentes electrónicos, como cámaras y microscopios de mayor resolución, ha mejorado significativamente la precisión y el reconocimiento de patrones en los softwares de detección precoz, haciéndolos altamente fiables.

La visión por computador sigue expandiéndose y se prevé que su impacto continuará creciendo en los próximos años, impulsando la automatización y la transformación digital en múltiples industrias.

DESAFÍOS DE LA VISIÓN POR COMPUTADOR

En visión por computador, la exactitud de los modelos depende en gran medida de la calidad y diversidad de los datos utilizados para entrenarlos. Sin embargo, los sesgos en los datos pueden causar problemas graves, como

resultados discriminatorios o incorrectos. Por ejemplo, si los datos de entrenamiento no incluyen suficiente diversidad en términos de género, etnia o condiciones ambientales, el modelo puede ser menos preciso al reconocer ciertas características. Este sesgo afecta la capacidad del sistema para generalizar, lo que es un desafío clave en aplicaciones críticas como seguridad y reconocimiento facial⁴.

La necesidad de hardware potente es un desafío técnico relevante en visión por computador. Los algoritmos de procesamiento de imágenes y el entrenamiento de modelos de aprendizaje profundo requieren un uso intensivo de recursos computacionales, como GPU, lo que puede resultar costoso y demandante en términos de energía. Además, la gestión de grandes volúmenes de datos es crítica; la recolección y almacenamiento de imágenes y videos requieren soluciones eficientes para asegurar la calidad de los datos, así como un procesamiento y preprocesamiento adecuados para evitar resultados inexactos. Sin la mejora de estos factores, el acceso y la escalabilidad de las soluciones en visión por computador serán limitadas.

FUTURO DE LA VISIÓN POR COMPUTADOR

Las nuevas tendencias en visión por computador se centran en la combinación con inteligencia artificial y aprendizaje automático avanzado. El aprendizaje auto-supervisado permite a los modelos aprender datos sin necesidad de que estos estén etiquetados previamente, mejorando de esta manera su eficiencia. Además, el uso de redes neuronales profundas está optimizando la visión 3D, crucial para el campo de la robótica y los vehículos autónomos.

Otro campo en el que se están consiguiendo avances considerables, es en la visión multimodal. La visión multimodal combina múltiples fuentes de datos sensoriales, como imágenes, texto, audio o video, para crear una comprensión más rica y contextualizada del entorno. Estos sistemas multimodales, no solo procesan imágenes, sino que, a su vez integran otras modalidades como descripciones textuales o señales sonoras, mejorando la precisión de

4 <https://confilegal.com/20240616-opinion-aprendizaje-automatico-y-gestion-de-datos-pilares-esenciales-para-el-funcionamiento-de-la-inteligencia-artificial/>

tareas complejas como la interpretación de escenas, análisis de sentimientos o reconocimiento de objetos. Esto permite aplicaciones como la realidad aumentada, en la que la visión por computador se complementa con información textual, o la traducción automática de video, donde el sistema puede reconocer tanto lo que se ve como lo que se escucha.

En la actualidad, la visión por computador en smartphones permite aplicaciones avanzadas como el reconocimiento facial, la fotografía computacional y la realidad aumentada. De cara al futuro, esta tecnología se encamina hacia una mayor integración de la realidad aumentada inmersiva, interacciones visuales avanzadas y el análisis de escenas en tiempo real. En drones, conseguirá mejorar la navegación autónoma en entornos complejos, permitirá el mapeo 3D detallado a gran escala y en tiempo real y la detección de obstáculos, haciéndolos más precisos y eficientes en tareas como la entrega de paquetes o la vigilancia aérea. Estos avances tanto en smartphones, como en drones, combinan sensores potentes con algoritmos de aprendizaje profundo, optimizados para hardware limitado, expandiendo sus capacidades en múltiples industrias.

En cuanto a los dilemas éticos relacionados con la vigilancia, el uso de datos visuales y el control de las tecnologías con inteligencia artificial, en marzo del presente año se publicó el siguiente reglamento: "Reglamento (UE) 2024/1689" en el que se establece un marco legal uniforme para la inteligencia artificial (IA) en la Unión Europea (UE), incluyendo aplicaciones de visión por computador. Esta normativa tiene como objetivo principal asegurar el desarrollo, el uso seguro y ético de la IA, promoviendo sistemas centrados en el ser humano. Entre sus puntos clave, se prohíben ciertas prácticas de IA, se establecen requisitos específicos para sistemas de alto riesgo, y se fomentan innovaciones en IA, particularmente para pymes y startups. Desde agosto de 2024, esta ley ha entrado en vigor y se implementará gradualmente para regular aplicaciones como el reconocimiento facial y otras tecnologías de visión por computador basadas en inteligencia artificial en los próximos años. Este proceso garantizará un desarrollo ético y seguro, enfocándose en la protección de derechos fundamentales y estableciendo normativas específicas para los sistemas de alto riesgo, como los relacionados con la categorización biométrica y la seguridad pública. 

LOS ROBOTS Y LA IA

💬 Ignacio Despujol Zabala

UN CAMINO QUE ESTÁ EMPEZANDO

Según la Real Academia Española un robot es “una máquina o ingenio electrónico programable que es capaz de manipular objetos y realizar diversas operaciones”. No hay un consenso sobre la definición exacta de robot, de hecho, los artículos de Wikipedia en español y en inglés amplían esta definición en diferentes dimensiones sin coincidir. Merece la pena incluir ambas definiciones porque capturan matices distintos e interesantes.

La Wikipedia en español dice “Un robot es una entidad virtual o mecánica artificial. En la práctica, esto es, por lo general, un sistema electromecánico que, por su apariencia o sus movimientos, ofrece la sensación de tener un propósito propio. La independencia en sus acciones hace que sus acciones sean la razón de un estudio razonable y profundo en el área de la ciencia y tecnología. La palabra puede referirse tanto a mecanismos físicos como a sistemas virtuales de software, aunque suele aludirse a los segundos con el término de bots.

No hay un consenso sobre qué máquinas pueden ser consideradas robots, pero sí existe un acuerdo general entre los expertos y el público sobre que los robots tienden a hacer parte o todo lo que sigue: moverse, hacer funcionar un brazo mecánico, sentir y manipular su entorno y mostrar un comportamiento inteligente, especialmente si ese comportamiento imita al de los humanos o a otros animales. Actualmente podría considerarse que un robot es una computadora con la capacidad y el propósito de movimiento que, en general, es capaz de desarrollar múltiples tareas de manera flexible según su programación; así que podría diferenciarse de algún electrodoméstico específico.”

La Wikipedia en inglés dice “Un robot es una máquina, especialmente una programable por computadora, capaz de llevar a cabo una serie compleja de acciones de forma automática. Un robot puede ser guiado por un dispositivo de control externo, o el control puede estar integrado en su interior. Los

robots pueden construirse para evocar la forma humana, pero la mayoría de los robots son máquinas que realizan tareas, diseñadas con un énfasis en la funcionalidad austera, en lugar de una estética expresiva. Los robots pueden ser autónomos o semiautónomos."

Como se puede ver, las definiciones coinciden en hablar de una máquina programable que interactúa con el entorno físico que le rodea realizando acciones por su cuenta de forma más o menos autónoma y programable (aunque en alguna definición se habla de que pueden ser guiados por un dispositivo externo, que puede ser controlado por un humano en lugar de programable y en alguna otra se habla de que pueden ser sistemas virtuales de software sin interacción con el entorno físico).

Hay robots de muchos tipos y para aplicaciones muy dispares. Sin entrar en si los sistemas virtuales de software son robots o no, podemos encontrar, entre muchos otros, robots humanoides, como el ASIMO de Honda, el Robot TOPIO Jugador de Ping Pong de TOSY o el más reciente Atlas de Boston Dynamics, el robot Perseverance que se mueve por Marte de forma autónoma, robots industriales en las cadenas de montaje (son muy conocidos los de las fábricas de coches), robots quirúrgicos, robots de asistencia al paciente, robots domésticos para la limpieza y mantenimiento del hogar, perros robot para el apoyo a la infantería militar, robots tractores y drones autónomos de fumigación y monitorización para la agricultura, robots de enjambre programados de forma colectiva para tareas colaborativas, drones UAV como el MQ-1 Predator de General Atomics e incluso nano robots microscópicos. Hay una guía muy buena online en inglés sobre robots: <https://robotsguide.com>



LA IA Y LOS ROBOTS

Los robots tienen sensores para percibir el entorno, un sistema de decisión para decidir qué hacer en base a la información que reciben y actuadores para interactuar con él.

Los sistemas de decisión de los robots van desde un simple sistema de control remoto manejado por el ser humano, como en la mayoría de los

robots quirúrgicos de primera generación (los más modernos ya incorporan asistencia con IA), los robots artificiales o en drones como el MQ-1 Predator (aunque hay drones totalmente autónomos como los drones bomba iraníes usados por Rusia en la guerra de Ucrania), y que no encaja en algunas de las definiciones de robot al no ser automático, hasta sistemas completamente autónomos basados en multitud de sensores y redes neuronales profundas como los utilizados en los coches autónomos.

Entre ambos hay un sinfín de mecanismos de control que van incrementando la autonomía de los robots, desde los más simples como los de los robots industriales clásicos de la industria del automóvil, basados en posiciones y tiempos coordinados para realizar las acciones, pasando a los que incorporan a sus decisiones la información de sensores externos, como los robots de clasificación de fruta por visión o los robots de limpieza que mapean su entorno con diversos sensores y crean una ruta en función del mapa obtenido, hasta los coches totalmente autónomos.

Aunque en los mecanismos más simples no se puede hablar de Inteligencia Artificial, conforme se avanza en complejidad de control y se gana en autonomía, se va incorporando la inteligencia artificial a los sistemas de decisión, desde sistemas de visión artificial y aprendizaje automático sencillos hasta técnicas mucho más avanzadas como las redes neuronales profundas y sistemas de aprendizaje por refuerzo que incorporan los últimos avances en el campo de la IA.

EL PROBLEMA DE LA DISPONIBILIDAD DE DATOS

La revolución de la IA generativa, representada por herramientas como Chat-GPT, DALL-E o Midjourney, se basa en una fórmula sencilla: entrenar una red neuronal grande con enormes conjuntos de datos extraídos de la web para cumplir con solicitudes dispares de los usuarios. Los modelos de lenguaje como estos pueden responder preguntas, escribir código o crear poesía, mientras que los sistemas de generación de imágenes pueden producir arte convincente a partir de una descripción.

Sin embargo, estas capacidades sorprendentes de la IA no se han traducido en robots como los que vemos en la ciencia ficción, capaces de limpiar, doblar ropa o preparar el desayuno. Esto se debe a que la fórmula de IA

generativa, basada en grandes modelos entrenados con datos de Internet, no es fácil de aplicar a la robótica. La web está llena de datos de texto e imágenes, pero no de interacciones robóticas, y las interacciones con el mundo real (como coger una taza de una mesa), aunque sean sencillas para nosotros, son, en la mayoría de los casos, complejas y están llenas de incertidumbres (que los humanos resolvemos de forma automática, generalizando el conocimiento que hemos aprendido en un dominio para usarlo en otro y aplicando el "sentido común", todo sin darnos cuenta). Los robots necesitan datos específicos de robótica para aprender, y estos datos se generan de forma lenta y laboriosa en entornos de laboratorio para tareas muy concretas. A pesar del progreso en los algoritmos de aprendizaje para robots, sin datos suficientes, todavía no es posible que los robots realicen tareas generales del mundo real fuera del laboratorio como las que llevamos a cabo en una casa o en una oficina.

Ya hay esfuerzos conjuntos de laboratorios de robótica de todo el mundo para crear grandes conjuntos de datos con datos de interacciones robóticas y combinarlos con los modelos de razonamiento semántico a partir de imágenes ya existentes en Internet para que puedan utilizarse para hacer robots de uso general, capaces de responder a instrucciones como "Pon la manzana entre la naranja y el plato" (<https://spectrum.ieee.org/global-robotic-brain>).



También se está estudiando el uso de técnicas de aprendizaje de transferencia. Esta metodología permite que los sistemas de IA tomen lo que han aprendido en un dominio y lo apliquen a otro, reduciendo la necesidad de recopilar grandes cantidades de datos en cada nuevo entorno. Por ejemplo, un robot que ha aprendido a identificar ciertos objetos en una fábrica podría transferir parte de ese conocimiento a un nuevo entorno sin tener que empezar desde cero.

Además, los desarrollos en algoritmos de aprendizaje por refuerzo han abierto nuevas posibilidades. Este tipo de aprendizaje permite que los robots mejoren a través de prueba y error en lugar de depender únicamente de datos preexistentes. Un robot que aprende a moverse en un entorno

desconocido, por ejemplo, puede refinar sus movimientos basándose en la retroalimentación de sus propias acciones.

Otra aproximación es crear datos sintéticos de interacciones robóticas a partir de simulaciones a las que incorporan aleatoriedad para simular el mundo real y utilizar estos datos para entrenar los sistemas. Por ejemplo, empresas como OpenAI utilizan simulaciones para entrenar a robots en tareas como el montaje de objetos o la navegación en espacios tridimensionales (<https://spectrum.ieee.org/openai-demonstrates-complex-manipulation-transfer-from-simulation-to-real-world>).



Estas técnicas y su combinación están produciendo ya avances significativos en el mundo de la robótica, que se van incorporando en todos los ámbitos de este campo, pero todavía estamos lejos de crear una inteligencia robótica lo suficientemente general como para realizar de forma fiable las tareas que realizamos los humanos en el día a día de nuestras vidas.

DESAFÍOS ÉTICOS Y SOCIALES

La integración de la IA y la robótica plantea importantes dudas éticas y sociales. Uno de los principales desafíos es el impacto de la automatización en el empleo. En la manufactura, el transporte y otros sectores, la automatización ha comenzado a reemplazar trabajos que anteriormente realizaban humanos. Aunque algunos argumentan que estos avances crearán nuevos tipos de empleo, otros temen que la tasa de destrucción de empleos supere la creación de nuevas oportunidades.

Otro problema ético clave es la cuestión de la responsabilidad. Si un robot comete un error que resulta en daño, ¿quién es el responsable? Por ejemplo, en el caso de los autos autónomos, ¿debería ser el fabricante del vehículo, el desarrollador del software o el propio usuario quien asuma la responsabilidad en caso de un accidente?

Además, la integración de IA en sistemas robóticos plantea preocupaciones sobre la privacidad y la seguridad. A medida que los robots se vuelven más inteligentes y están más conectados, existe el riesgo de que sean hackeados o utilizados de maneras que comprometan la privacidad de las personas.

Y la última pregunta tiene que ver con cuánta autonomía deberíamos darles a unos sistemas con capacidad de actuación que pueden manejar sistemas críticos y cómo mantener el control humano sobre estas tecnologías para evitar un hipotético efecto "Terminator".

EL FUTURO DE LA IA Y LA ROBÓTICA

A pesar de los desafíos, el futuro de la IA y la robótica parece brillante. Los avances en algoritmos y el aumento en la capacidad de procesamiento están permitiendo desarrollar robots cada vez más sofisticados. En los próximos años iremos viendo cada vez más robots en sectores como la agricultura, la construcción, la medicina y los servicios públicos.

Además, a medida que se resuelvan los problemas relacionados con la disponibilidad de datos y se desarrollen mejores marcos éticos, la integración de la IA en la robótica permitirá crear robots más adaptables y versátiles que tendrán un impacto más profundo en la sociedad.

En el futuro cercano no veremos robots ayudándonos en casa como en las películas de ciencia ficción, pero sí se irán incorporando más robots y cada vez más autónomos a la realización de distintas tareas en cada vez más sectores, mejorando de forma significativa su productividad y eficiencia. 📄

SISTEMAS EXPERTOS

💬 Jesús Alfaro, Armando Ortuño y Enric Montesa

NACIMIENTO DE LOS SISTEMAS EXPERTOS

Los *sistemas expertos* son una rama pionera de la inteligencia artificial (IA) que surgió en la década de 1960, cuando los investigadores comenzaron a explorar la posibilidad de crear programas capaces de replicar el razonamiento humano experto en dominios específicos. Uno de los primeros hitos fue **DENDRAL** (1965), un sistema diseñado para inferir la estructura molecular de compuestos químicos a partir de datos espectrométricos. **DENDRAL** fue desarrollado por Edward Feigenbaum, considerado el “*padre*” de los sistemas expertos, junto con Joshua Lederberg y Bruce Buchanan.

A partir de **DENDRAL**, se construyeron sistemas como **MYCIN** (1972), que fue un hito en el campo de la medicina, utilizado para diagnosticar infecciones bacterianas y recomendar tratamientos. Aunque nunca fue implementado en hospitales, **MYCIN** demostró el potencial de los sistemas expertos para asistir en decisiones complejas basadas en grandes volúmenes de conocimiento.

EVOLUCIÓN Y HITOS PRINCIPALES

Durante las décadas de 1970 y 1980, los sistemas expertos ganaron popularidad debido a su capacidad de replicar el conocimiento de expertos humanos en campos específicos como la medicina, la ingeniería, la química y las finanzas. Un sistema experto típico se componía de tres partes principales:

1. **Base de conocimiento:** una colección de hechos y reglas sobre un dominio específico.
2. **Motor de inferencia:** un conjunto de algoritmos que usaba las reglas y hechos de la base de conocimiento para realizar deducciones y resolver problemas.
3. **Interfaz de usuario:** la parte que permitía a los usuarios interactuar con el sistema, ingresando información y recibiendo recomendaciones.

Uno de los hitos más importantes fue el desarrollo de **XCON** en 1980, un sistema experto utilizado por la empresa *Digital Equipment Corporation* (DEC) para configurar equipos informáticos de manera eficiente. **XCON** ahorró millones de dólares a la empresa, lo que generó un gran interés en los sistemas expertos como herramientas para la toma de decisiones.

PROMESAS Y EXPECTATIVAS

En sus primeros días, los sistemas expertos generaron grandes expectativas. Se pensaba que podrían replicar, e incluso superar, la capacidad de los humanos en tareas especializadas, democratizando el acceso a conocimientos altamente técnicos. Se prometía que estos sistemas serían capaces de:

- Tomar decisiones complejas en áreas críticas como la medicina y la ingeniería.
- Reducir el tiempo y costo de consultoría especializada.
- Resolver problemas en tiempo real en sectores como el comercio, la manufactura y la atención médica.

La capacidad de los sistemas expertos para manejar grandes volúmenes de datos y aplicar reglas lógicas parecía una solución ideal para enfrentar la creciente complejidad de las industrias modernas.

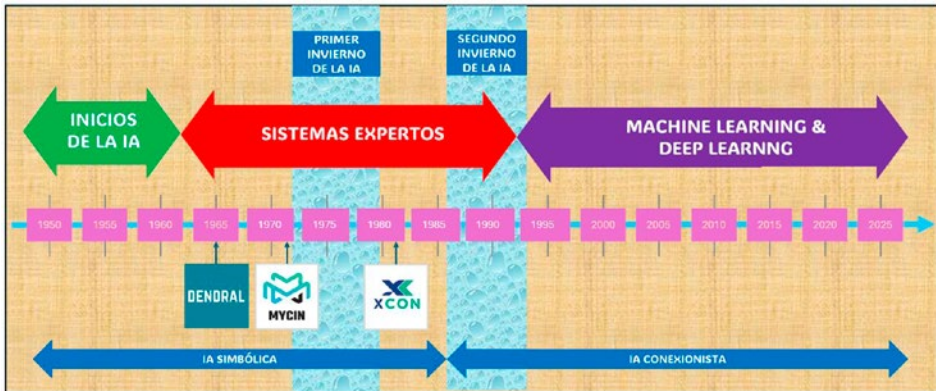
FRACASOS Y LIMITACIONES

Como ya se ha comentado en un capítulo anterior, a pesar de los avances, los sistemas expertos comenzaron a encontrar varias limitaciones:

- Dependencia de expertos humanos: La creación de una base de conocimiento requería que expertos humanos codificaran sus conocimientos, lo que resultaba costoso y lento.
- Dominio limitado: Estos sistemas eran muy efectivos en dominios específicos, pero no podían transferir su conocimiento a otras áreas. MYCIN, por ejemplo, era excelente en diagnóstico de infecciones bacterianas, pero no podía aplicarse a otras ramas de la medicina.

- **Mantenimiento costoso:** A medida que el conocimiento avanzaba, actualizar las reglas y la base de conocimiento se convertía en una tarea compleja, costosa y poco escalable.
- **Rigidez:** Los sistemas expertos seguían reglas predefinidas, por lo que tenían dificultades para manejar situaciones inesperadas o dinámicas en las que el conocimiento podía cambiar rápidamente.

Estas limitaciones llevaron a una disminución en el interés por los sistemas expertos a fines de la década de 1980 y principios de 1990, siendo conocido el periodo 1987-1993 como el segundo invierno de la IA. El declive de los sistemas expertos coincidió con el auge de otros enfoques en IA, como las redes neuronales (ANN) y el aprendizaje automático (Machine Learning: ML).



RELACIÓN DE LOS SISTEMAS EXPERTOS CON LA GESTIÓN DEL CONOCIMIENTO

Los *sistemas expertos* y la *gestión del conocimiento* están estrechamente relacionados, ya que ambos tratan de capturar, organizar y utilizar el conocimiento de manera efectiva para resolver problemas o mejorar procesos. Para entender esta relación, es útil explorar cómo los sistemas expertos contribuyen al campo de la gestión del conocimiento y cómo evolucionaron en paralelo.

Captura y Formalización del Conocimiento

Uno de los objetivos fundamentales de los sistemas expertos es la *captura del conocimiento* especializado de los seres humanos y su traducción en una

forma que las máquinas puedan utilizar. Este proceso de formalización del conocimiento se realiza mediante la creación de una *base de conocimiento* en la que se almacenan hechos, reglas y procedimientos obtenidos de expertos humanos. Este es un elemento central en la *gestión del conocimiento*, que busca capturar y organizar el conocimiento tácito (aquellos que los expertos poseen, pero no está documentado) y convertirlo en conocimiento explícito (formalizado y accesible).

Por ejemplo, en un entorno de gestión del conocimiento, un sistema experto puede ayudar a documentar los procedimientos y decisiones que toma un ingeniero o un médico, para que este conocimiento pueda ser reutilizado por otros profesionales, acelerando la toma de decisiones y evitando la pérdida de conocimiento crítico cuando un experto deja una organización.

Almacenamiento y Acceso al Conocimiento

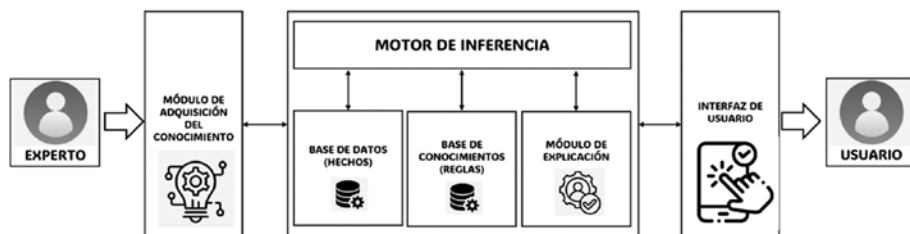
Los sistemas expertos proporcionan un medio para almacenar grandes cantidades de conocimiento en un formato estructurado, lo que es esencial para la gestión del conocimiento. En lugar de tener que depender de la disponibilidad continua de un experto humano, un sistema experto permite a los usuarios acceder al conocimiento almacenado en cualquier momento. Esto se alinea con la gestión del conocimiento, que tiene como objetivo crear repositorios accesibles donde el conocimiento crítico pueda ser consultado fácilmente por quienes lo necesiten.

Este acceso a conocimiento formalizado permite una distribución más eficiente del conocimiento dentro de una organización, asegurando que no se concentre en unas pocas personas, sino que esté disponible de manera generalizada.

Aplicación del Conocimiento en la Toma de Decisiones

La *gestión del conocimiento* no se limita a la captura y almacenamiento de información, sino que también busca facilitar la *aplicación efectiva del conocimiento* para mejorar la toma de decisiones. Aquí es donde los sistemas expertos juegan un papel crucial. Estos sistemas no solo almacenan conocimiento, sino que lo utilizan activamente para resolver problemas, realizar diagnósticos o recomendar acciones.

El *motor de inferencia* de un sistema experto simula el proceso de razonamiento de un experto humano, aplicando las reglas almacenadas para generar soluciones o recomendaciones basadas en el conocimiento disponible. En un contexto de gestión del conocimiento, esto significa que el sistema puede ayudar a tomar decisiones más rápidas y precisas, al ofrecer soluciones basadas en el conocimiento experto previamente capturado.



Transferencia y Compartición del Conocimiento

Un desafío común en la gestión del conocimiento es la *transferencia efectiva del conocimiento* entre personas o dentro de una organización. Los sistemas expertos facilitan esta transferencia al encapsular el conocimiento de los expertos y ponerlo a disposición de otros usuarios, independientemente de su nivel de experiencia. Por ejemplo, un sistema experto en ingeniería puede ayudar a un ingeniero menos experimentado a resolver un problema complejo, guiándolo a través del proceso de razonamiento que seguiría un experto.

En términos más amplios, la gestión del conocimiento busca mejorar la *compartición del conocimiento* entre diferentes partes de una organización. Los sistemas expertos son una herramienta que contribuye a este objetivo, ya que permiten que el conocimiento de expertos en diferentes áreas se transfiera de manera estructurada y eficiente, sin necesidad de interacción directa.

Limitaciones y Desafíos en la Gestión del Conocimiento

A pesar de sus ventajas, los sistemas expertos también enfrentan limitaciones que repercuten en la gestión del conocimiento:

- **Conocimiento estático:** El conocimiento en los sistemas expertos tiende a ser estático, ya que se basa en reglas y hechos codificados en un momento determinado. En la gestión del conocimiento, el desafío es mantener este conocimiento actualizado y relevante en un entorno donde el conocimiento evoluciona rápidamente.
- **Dificultad en la captura del conocimiento tácito:** Los sistemas expertos son efectivos al capturar conocimiento explícito, pero tienen dificultades para capturar el conocimiento tácito, que a menudo es subjetivo y difícil de formalizar. La gestión del conocimiento enfrenta un desafío similar: cómo convertir el conocimiento tácito, basado en la experiencia y el juicio, en algo accesible y útil para otros.

Evolución hacia Nuevas Herramientas de Gestión del Conocimiento

Si bien los sistemas expertos fueron una de las primeras herramientas utilizadas para la gestión del conocimiento, hoy en día han sido complementados o reemplazados por tecnologías más avanzadas. *Sistemas de gestión del conocimiento* modernos, que incluyen bases de datos dinámicas, motores de búsqueda inteligentes y sistemas basados en aprendizaje automático, están diseñados para ser más flexibles y adaptativos que los sistemas expertos tradicionales. Estos nuevos sistemas no solo almacenan y aplican conocimiento, sino que también lo actualizan continuamente a medida que se dispone de nuevos datos.

A pesar de ello, la idea subyacente de capturar el conocimiento experto y formalizarlo para su uso automatizado sigue siendo un concepto fundamental, y los sistemas expertos han dejado un legado importante en este campo.

FLUJOGRAMA QUE VA DE LOS DATOS A LA ACCIÓN

El proceso que va desde la *captura de datos* hasta la *acción* incluye varias etapas clave que permiten transformar datos en decisiones fundamentadas. Este flujo es típico en procesos de toma de decisiones, tanto en sistemas automatizados como en entornos de gestión del conocimiento. A continuación, se comenta cada etapa de este proceso.

1. *Datos*

Esta es la fase inicial, en la que se recolectan los datos crudos o sin procesar a partir de diversas fuentes. Los datos pueden provenir de sensores, encuestas, bases de datos, transacciones financieras, dispositivos IoT, o cualquier otro sistema que genere información. En esta etapa, la precisión y cantidad de los datos son esenciales para que las siguientes etapas sean efectivas.

La captura de datos debe ser completa, precisa y relevante para el problema o contexto. Una mala captura puede llevar a decisiones erróneas más adelante. Además, hoy en día, con grandes volúmenes de datos (big data), la capacidad de capturar y almacenar información de manera eficiente es crucial.

2. *Información*

Una vez capturados los datos, estos deben ser *procesados* y convertidos en información significativa. Este proceso implica limpiar los datos, organizarlos y estructurarlos para que sean útiles. La *elaboración* puede incluir la agregación de datos, la normalización, la categorización o la transformación de formatos de datos brutos en algo interpretable.

La información es el resultado de procesar los datos crudos y darle contexto. En este punto, el uso de tecnologías como herramientas de procesamiento de datos o técnicas de análisis estadístico es crucial para extraer valor de los datos. Aquí es donde se eliminan errores, duplicados y se asegura la consistencia de la información.

3. *Conocimiento*

Esta etapa implica la interpretación y *comprensión* de la información procesada. En un contexto humano, es donde los responsables de la toma de decisiones analizan la información para extraer patrones, entender tendencias y contextualizar los datos. En un sistema automatizado, como un sistema experto o una IA, la asimilación se realiza mediante algoritmos que interpretan los resultados procesados. La asimilación o entendimiento de la es una fase crítica, ya que la correcta interpretación de la información define la calidad de las decisiones que se tomarán. Aquí se aplican

herramientas de análisis como cuadros de mando, informes o modelos predictivos. Además, los sistemas expertos y de IA modernos utilizan técnicas como el aprendizaje automático para asimilar grandes volúmenes de información y aprender de ellos.

4 . *Decisión*

Una vez asimilada la información, se pasa a la fase de ****toma de decisiones****. Esta etapa implica seleccionar la mejor opción o curso de acción basado en la información procesada. Dependiendo del contexto, la toma de decisiones puede ser automatizada (por un sistema experto o de IA) o realizada por humanos que utilizan la información disponible para evaluar alternativas.

La calidad de las decisiones está directamente ligada a la calidad de los datos y a la precisión de su interpretación. En esta fase, se suelen usar sistemas de soporte a la decisión (DSS), que ayudan a los tomadores de decisiones a considerar distintas alternativas y evaluar los riesgos y beneficios asociados a cada opción.

5 . *Acción*

Finalmente, la toma de decisiones lleva a una ****acción****, que es la ejecución de la opción seleccionada. Esta puede incluir la implementación de un plan, la aplicación de una política, la modificación de un proceso o cualquier otra intervención basada en la decisión tomada. En sistemas automatizados, la acción puede incluir cambios automáticos en un sistema o la activación de ciertos procedimientos.

La acción es el resultado final del proceso de decisión. Aquí es crucial medir el impacto de la acción y retroalimentar el sistema con nuevos datos para cerrar el ciclo de mejora continua. En muchas ocasiones, se requiere un



monitoreo constante para ajustar la acción según las variaciones en los datos o las nuevas circunstancias.

ESTADO ACTUAL

Como se expone en varios capítulos de la Guía, los sistemas expertos originales, han sido reemplazados en gran medida por enfoques más avanzados de inteligencia artificial. Sin embargo, muchos de sus principios fundamentales siguen siendo importantes y se han integrado en sistemas modernos.

- **Aprendizaje automático y Big Data:** A diferencia de los sistemas expertos, que dependían de reglas programadas, el aprendizaje automático permite a los sistemas aprender directamente de los datos. Esto elimina la necesidad de una codificación explícita de conocimientos y permite manejar grandes volúmenes de información con una mayor flexibilidad.
- **Sistemas híbridos:** Algunos sistemas actuales combinan el aprendizaje automático con reglas de sistemas expertos para maximizar la eficiencia y precisión. Por ejemplo, en la medicina, hay sistemas que usan algoritmos de aprendizaje profundo para analizar imágenes médicas, pero aplican reglas lógicas derivadas de expertos para hacer recomendaciones finales.
- **Asistentes inteligentes:** Aunque los sistemas expertos clásicos han disminuido, su legado puede verse en herramientas como asistentes virtuales o chatbots avanzados, que usan tanto aprendizaje automático como conocimiento estructurado para responder a preguntas y realizar tareas.

BURBUJA INVERSORA DE LOS AÑOS OCHENTA

Y EL PAPEL DE LA INVERSIÓN PÚBLICA Y PRIVADA

La década de los años 80 del siglo pasado, vivió una **burbuja de inversión** relacionada con la inteligencia artificial, impulsada en gran medida por las expectativas desmesuradas sobre los Sistemas Expertos. Varias empresas y gobiernos vieron en la IA la inminente gran revolución tecnológica lo que no fue otra cosa que una especie de "veranillo de la IA", un periodo de entusiasmo y optimismo particularmente en torno a los sistemas expertos que al igual que el veranillo de San Martín solo supuso un breve respiro de calor antes



del invierno, una época que trajo una oleada de inversiones y expectativas elevadas antes de un periodo más frío y realista en el desarrollo de la IA.

La fuente de las inversiones procedió tanto de **fondos públicos** como **privados**. En Estados Unidos, la DARPA (Agencia de Proyectos de Investigación Avanzados de Defensa) fue uno de los principales impulsores del desarrollo de IA, invirtiendo millones de dólares en proyectos relacionados con *Sistemas Expertos* para aplicaciones militares y de defensa. Otros gobiernos, como el japonés, lanzaron el famoso **Proyecto de la Quinta Generación**, una iniciativa ambiciosa para desarrollar computadoras que pudieran realizar tareas de razonamiento avanzado.

Tampoco hay que olvidar la apuesta realizada por las grandes corporaciones como **Xerox**, **IBM** y startups tecnológicas también invirtieron en la creación de *Sistemas Expertos* con la esperanza de revolucionar sectores como la medicina, la ingeniería y las finanzas. La financiación privada llegó en parte a través de venture capital (capital de riesgo), lo que influyó en la creación de múltiples startups enfocadas en IA.

Sin embargo, la tecnología de los *Sistemas Expertos* no cumplió con las expectativas. Las limitaciones de los algoritmos, la incapacidad de generalizar el conocimiento y la falta de potencia computacional para manejar problemas complejos llevaron al estallido de la **burbuja inversora** a finales de los años 80.

Los fondos públicos y privados empezaron a disminuir considerablemente, y muchos proyectos se cancelaron. Como ya se ha visto, este fenómeno fue conocido como el “**invierno de la IA**”, un periodo de estancamiento en la investigación y el desarrollo de inteligencia artificial que duró hasta mediados de la década de los 90. En definitiva, un período de altos y bajos en la inversión que sentó las bases para un entendimiento más realista del desarrollo de la IA, que con el tiempo ha evolucionado hacia enfoques más efectivos, como el *Aprendizaje Profundo* (Deep Learning).

HACIA UN NUEVO PARADIGMA: EL NEUROSIMBOLISMO

Los Sistemas Expertos juegan un papel crucial en la gestión del conocimiento al capturar, estructurar y aplicar el conocimiento especializado de manera automatizada. Estas herramientas permiten formalizar el conocimiento tácito de expertos humanos, convirtiéndolo en reglas y hechos almacenados en bases de conocimiento. Esto no solo preserva el saber experto, sino que lo hace accesible a otros usuarios y facilita su uso en diversos contextos, eliminando la dependencia de la disponibilidad de los expertos humanos para cada consulta.

En cuanto al apoyo a la toma de decisiones, los sistemas expertos son capaces de analizar información compleja y ofrecer recomendaciones basadas en el conocimiento almacenado. Al aplicar algoritmos a esta base de conocimiento a través de motores de inferencia, los sistemas pueden ofrecer soluciones rápidas y precisas en situaciones donde se requieren decisiones informadas, como en la medicina, la ingeniería o la gestión empresarial. De esta manera, se mejora la eficiencia y se reducen los errores al optimizar el proceso de decisión, brindando un soporte valioso en situaciones críticas.

Sin embargo, estos sistemas también enfrentan limitaciones, como la dificultad para adaptarse a contextos imprevistos o el mantenimiento constante que requieren para mantenerse actualizados. A pesar de ello, su integración con nuevas tecnologías de IA, como el Aprendizaje Automático, ha permitido superar algunas de estas barreras, lo que mantiene su relevancia en la gestión del conocimiento y el soporte a la toma de decisiones, justificando el nacimiento de lo que algunos ya se han atrevido a bautizar como la *IA neurosimbólica*, un enfoque que combina dos paradigmas tradicionales de

la inteligencia artificial: las redes neuronales (aprendizaje profundo) y el razonamiento simbólico.

Esta fusión busca aprovechar las fortalezas de ambos métodos para crear sistemas de IA más robustos y versátiles, lo cual permite:

- Abordar la falta de interpretabilidad de los modelos de Aprendizaje Profundo.
- Reducir la necesidad de grandes cantidades de datos para el aprendizaje.
- Mejorar la capacidad de capturar estructuras compositivas y causales en los datos. 📄

AGENTES INTELIGENTES Y LOS SISTEMAS MULTIAGENTE

💬 Vicent Botti

Si hablamos de Inteligencia Artificial (IA) tenemos que hablar del concepto de agente, ya que esta abstracción es, actualmente, fundamental para comprender lo que es la IA, es una abstracción utilizada ampliamente en la docencia de la IA pero también en la aplicación práctica de los sistemas de IA.

¿QUÉ ES UN AGENTE?

Si consultamos el Diccionario de la RAE encontramos como definición de agente:

“el que obra o tiene virtud de obrar, la persona que reproduce un efecto, la persona que obra en poder de otra, el que actúa en representación de otro (agente artístico, comercial, inmobiliario, de seguros, de bolsa, etc.)”.

Aunque parte de esta definición se ajusta en parte a lo que en IA entendemos como “agente”, este no es el concepto que tenemos de agente, nosotros estamos pensando en “agentes software”, de forma general en “aplicaciones informáticas con capacidad para decidir cómo deben actuar para alcanzar sus objetivos que pueden funcionar fiablemente en un entorno rápidamente cambiante e impredecible”.

El concepto de *agente inteligente* es un concepto que nace del área de la inteligencia artificial; de hecho, una definición comúnmente aceptada relaciona la disciplina de la inteligencia artificial con el *análisis y diseño de entidades autónomas capaces de exhibir un comportamiento inteligente*. Un agente autónomo es un sistema o entidad que tiene la capacidad de operar de manera independiente, sin intervención directa de un humano o de otros sistemas, para lograr ciertos objetivos. Estos agentes están equipados con sensores para percibir su entorno, y con actuadores o mecanismos para tomar acciones que afecten ese entorno. La autonomía del agente radica en su *capacidad para tomar decisiones por sí mismo* basándose en la información que recoge, procesarla y adaptarse a cambios en el entorno sin depender de un control externo constante.

Una definición ampliamente del concepto de agente es la proporcionada por *Michael Wooldridge*¹:

“Cualquier proceso computacional dirigido por el objetivo capaz de interactuar con su entorno de forma flexible y robusta”.

Dicho en otras palabras, un agente es un sistema informático, situado en algún entorno, que *percibe el entorno* (entradas sensibles de su entorno) y a partir de tales percepciones *determina* (mediante técnicas de resolución de problemas) *y ejecuta acciones* (de forma autónoma y flexible) *que le permiten alcanzar sus objetivos* y que pueden cambiar el entorno.

En esta definición de agente se incluyen una serie de propiedades que un software de satisfacer para que, al menos en el área de IA, lo consideremos un agente:

Reactividad: Sistema reactivo mantiene una constante interacción con su entorno, percibe el entorno y responde/actúa (a tiempo para que la respuesta sea útil) a los cambios que ocurren en él.

Proactividad/Iniciativa: es la capacidad del agente generar e intentar conseguir objetivos, el agente no está dirigido solamente por eventos, tiene capacidad de tomar la iniciativa.

Sociabilidad: es la capacidad de interactuar con otros agentes (y posiblemente con humanos) mediante cooperación, coordinación y negociación, para lo que el agente requiere la capacidad de comunicarse con otros agentes.

¿QUÉ ES UN SISTEMA MULTIAGENTE (SMA)?

El mundo real es un mundo multiagente: para resolver muchos problemas del mundo real, cuando queremos conseguir objetivos debemos tener en cuenta al resto de entidades (agentes) del entorno donde nos encontramos. En algunos casos para alcanzar un objetivo es necesario interactuar con otros ya que puede que nosotros no tengamos capacidades suficientes para lograrlo.

1 Wooldridge, M. (2009). *An Introduction to MultiAgent Systems* (2nd ed.). Chichester, England: John Wiley & Sons. ISBN 978-0-470-51946-2

Un *sistema multiagente* está constituido por un conjunto de agentes (dos o más agentes) que interactúan los unos con los otros. En el caso más general, los agentes actuarán en representación de usuarios que tienen diferentes objetivos y motivaciones. Para interactuar con éxito, requerirán capacidades para cooperar, coordinarse y negociar con cada uno de los otros, tal como hace la gente.

Los sistemas multiagente son un caso particular de un sistema distribuido, y su particularidad radica en el hecho de que los componentes del sistema son autónomos y egoístas, buscando satisfacer sus propios objetivos. Además, estos sistemas también se destacan por ser sistemas abiertos sin un diseño centralizado. Una de las principales razones del gran interés y atención que han recibido los sistemas multiagente es que se ven como una tecnología habilitadora para aplicaciones complejas que requieren procesamiento distribuido y paralelo de datos, y que operan de manera autónoma en dominios complejos y dinámicos.

La investigación en la disciplina de los *sistemas multiagente (SMA)* se basa en los resultados de la computación distribuida, formulando nuevas preguntas sobre cómo los agentes deben interactuar entre sí para coordinar sus actividades y resolver problemas complejos. La mayor parte de la investigación actual se centra en diseñar mecanismos de coordinación adecuados para gestionar coaliciones o equipos de agentes. La programación de agentes inteligentes plantea desafíos complejos para los ingenieros, porque además de la complejidad de diseñar sistemas concurrentes y distribuidos, se añade la complejidad de que los componentes deben tener una arquitectura que incluya aspectos como reactividad, proactividad y sociabilidad. Estas propiedades no son fáciles de programar cuando el entorno es dinámico y complejo. Para lograr una programación real de agentes, se han hecho numerosas propuestas por parte de muchos investigadores, tanto para arquitecturas de agentes, como para lenguajes de comunicación y mecanismos de toma de decisiones y coordinación. En este último caso, surge la necesidad de que los agentes puedan llegar a acuerdos para operar en un sistema multiagente. Aquí radica un aspecto importante de la complejidad de la programación de agentes inteligentes.

Características principales de un sistema multiagente son:

1. **Descentralización:** No existe una autoridad central que controle a todos los agentes. En su lugar, cada agente toma decisiones localmente y de manera autónoma en función de la información que posee.
2. **Interacción:** Los agentes pueden comunicarse y colaborar entre sí o competir para alcanzar sus objetivos. Las interacciones entre ellos pueden ser directas (a través de mensajes) o indirectas (mediante cambios en el entorno compartido).
3. **Heterogeneidad:** Los agentes pueden ser diferentes entre sí en términos de capacidades, conocimiento, comportamientos o funciones dentro del sistema. Algunos pueden ser simples, mientras que otros pueden ser más complejos o especializados.
4. **Cooperación o Competencia:** Dependiendo del diseño del sistema, los agentes pueden cooperar para alcanzar un objetivo común o competir si tienen objetivos en conflicto.
5. **Autonomía:** Cada agente toma decisiones de forma independiente, basándose en su propio conocimiento y en la información que recopila del entorno o de otros agentes.
6. **Distribución de conocimiento:** Ningún agente tiene acceso a toda la información del sistema. Los agentes dependen de la información local o compartida con otros para tomar decisiones.

En el marco de la inteligencia artificial, los sistemas multiagente se han caracterizado por ofrecer una posible solución al desarrollo de problemas complejos con características distribuidas. Al abordar el desarrollo de sistemas multiagente, sin duda hay un aumento significativo en la complejidad, así como la necesidad de adaptar técnicas existentes, o en algunas situaciones, desarrollar nuevas técnicas y herramientas. En los últimos años, han aparecido diferentes trabajos que intentan proponer nuevos procesos y técnicas para el desarrollo de sistemas multiagente.

Algunas de las principales aplicaciones de los sistemas multiagente son:

- **Entornos industriales** (robótica colaborativa, optimización de procesos, cadenas de suministro, ...)

- **Simulación social y económica** (modelado de sociedades humanas, simulaciones de organizaciones, organizaciones virtuales, ...)
- **Gestión de tráfico y movilidad urbanas** (sistemas de control de tráfico, reparto de última milla, ...)
- **Sistemas financieros y mercados electrónicos, comercio electrónico** (mercados bursátiles o de valores, gestión de riesgo, recomendadores, asistentes personalizados, ...)
- **Sistemas de energía y redes inteligentes** (optimización de consumo eléctrico, optimización de la distribución de energía, ...)
- **Ciudades Inteligentes** (gestión de las redes de distribución de agua potable o de aguas residuales, planificación urbanística, ...)
- **Salud** (asistentes personales telemedicina, gestión de recursos hospitalarios, ...)

La construcción de **SMA** integra tecnologías de diferentes áreas del conocimiento: técnicas de ingeniería de software para estructurar el proceso de desarrollo; técnicas de IA para dotar a los sistemas de la capacidad de enfrentarse a situaciones inesperadas y tomar decisiones, y programación concurrente para abordar la coordinación de tareas ejecutadas en diferentes máquinas bajo diferentes políticas de planificación. Debido a esta combinación de tecnologías, el desarrollo de **SMA** es complicado. En este sentido, durante los últimos años, han aparecido diferentes plataformas y herramientas de desarrollo que ofrecen soluciones parciales para el modelado y diseño de sistemas basados en agentes.

Entre ellas podríamos destacar:

- JADE (Java -Agent Development Framework-)
- Spade
- NetLogo
- Repast
- Jason
- AgentScape
- Flame 

1. INTRODUCCIÓN

La inteligencia artificial ha revolucionado el panorama de la ciberseguridad, proporcionando capacidades avanzadas para detectar, prevenir y responder a amenazas ciber cada vez más sofisticadas. Su uso permite el análisis de grandes volúmenes de datos en tiempo real, la identificación de patrones complejos y la automatización de tareas repetitivas, lo que mejora significativamente la eficacia y eficiencia de las operaciones de seguridad.

Sus principales aplicaciones en ciberseguridad incluyen:

- Detección de amenazas avanzadas
- Análisis de comportamiento de usuarios y entidades (UEBA)
- Respuesta automatizada a incidentes
- Gestión y análisis de vulnerabilidades
- Prevención de fraudes y detección de anomalías
- Fortalecimiento de la autenticación y control de acceso

En el ámbito de la ciberseguridad, existen diferentes equipos identificados por colores que desempeñan roles específicos para proteger y evaluar la seguridad de las organizaciones:

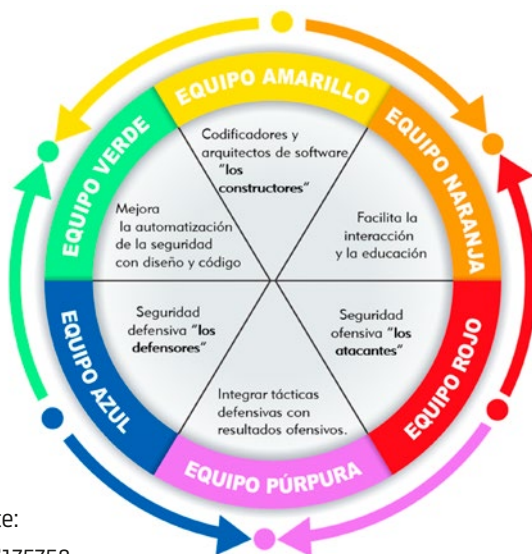


Figura 1: Diferentes equipos (Teams) en Ciberseguridad y sus roles. Fuente: https://x.com/isec_pe/status/1453759019798175753

Examinemos cómo cada equipo de seguridad aprovecha la IA para mejorar su eficiencia frente a los desafíos de ciberseguridad actuales.

EQUIPO AZUL (BLUE TEAM):

SEGURIDAD DEFENSIVA “LOS DEFENSORES”

El Blue Team en ciberseguridad es un equipo especializado de profesionales encargados de proteger y defender los sistemas, redes y activos digitales de una organización contra amenazas ciber. Sus principales funciones incluyen la implementación de medidas de seguridad, la monitorización continua de la infraestructura tecnológica, la detección temprana de intrusiones, la respuesta a incidentes y la realización de evaluaciones de riesgos. Este equipo trabaja de manera proactiva para fortalecer las defensas de la organización, establecer políticas de seguridad, educar a los empleados sobre mejores prácticas y utilizar herramientas avanzadas para proteger los activos críticos contra posibles ataques.

A continuación, se detalla como puede utilizar la Inteligencia Artificial (IA) en la mejora de sus capacidades defensivas.

Detección de amenazas en tiempo real

Para implementar sistemas de detección de amenazas más sofisticados y eficaces. Los algoritmos de aprendizaje automático pueden analizar el tráfico de red, los logs de sistemas y las actividades de usuarios en tiempo real para identificar patrones sospechosos o anomalías que podrían indicar un ataque en curso.

CASO PRÁCTICO: Implementación de un SIEM (Security Information and Event Management) potenciado por IA.

1. **Recolección de datos:** Configurar la importación de logs y eventos de múltiples fuentes (firewalls, servidores, aplicaciones, etc.).
2. **Entrenamiento del modelo:** Utilizar datos históricos para entrenar modelos de aprendizaje automático que puedan distinguir entre actividad normal y sospechosa.
3. **Análisis en tiempo real:** Procesar los datos entrantes a través de los modelos entrenados para detectar anomalías.

4. **Generación de alertas:** Configurar el sistema para generar alertas automáticas cuando se detecten actividades sospechosas.
5. **Mejora continua:** Retroalimentar el sistema con los resultados de las investigaciones para mejorar la precisión del modelo.

Respuesta automatizada a incidentes

Para automatizar las respuestas iniciales a ciertos tipos de incidentes, reduciendo el tiempo de reacción y mitigando el impacto potencial de las amenazas.

CASO PRÁCTICO: Implementación de un sistema de respuesta automatizada.

1. **Definición de escenarios:** Identificar los tipos de incidentes que pueden ser manejados de forma automatizada.
2. **Diseño de playbooks:** Crear flujos de trabajo detallados para cada escenario.
3. **Integración con sistemas:** Conectar el sistema de respuesta automatizada con firewalls, sistemas de control de acceso, etc.
4. **Implementación de IA:** Utilizar algoritmos de aprendizaje automático para mejorar la toma de decisiones en la respuesta a incidentes.
5. **Monitorización y ajuste:** Supervisar el rendimiento del sistema y ajustar los modelos según sea necesario.

Análisis predictivo de vulnerabilidades

Para predecir posibles vulnerabilidades antes de que sean explotadas, permitiendo una gestión proactiva de la seguridad.

CASO PRÁCTICO: Desarrollo de un sistema de análisis predictivo de vulnerabilidades.

1. **Recopilación de datos:** Agregar información sobre vulnerabilidades conocidas, parches, configuraciones de sistemas y tendencias de amenazas.
2. **Modelado predictivo:** Utilizar técnicas de aprendizaje automático para crear modelos que puedan predecir la probabilidad de nuevas vulnerabilidades.
3. **Integración con procesos:** Incorporar los resultados del análisis predictivo en los procesos de gestión de parches y configuración.
4. **Validación y refinamiento:** Comparar las predicciones con las vulnerabilidades reales descubiertas y ajustar los modelos en consecuencia.

EQUIPO VERDE (GREEN TEAM):

MEJORA DE LA AUTOMATIZACIÓN DE LA SEGURIDAD CON DISEÑO Y CÓDIGO

El Green Team es el equipo dedicado a la automatización en ciberseguridad implica el uso de herramientas, scripts y soluciones tecnológicas para realizar tareas repetitivas, monitorear y responder a eventos de seguridad, y agilizar los procesos de detección y respuesta ante amenazas.

A continuación, se detalla cómo puede beneficiarse significativamente de la IA para mejorar sus procesos y resultados:

Análisis de código seguro

Para identificar vulnerabilidades y problemas de seguridad en el código fuente durante el proceso de desarrollo.

CASO PRÁCTICO: Implementación de un sistema de análisis de código seguro con IA.

1. **Selección de herramientas:** Elegir una plataforma de análisis de código estático (SAST) que incorpore capacidades de IA.
2. **Integración en el ciclo de desarrollo:** Configurar la herramienta para que se ejecute automáticamente en cada guardado (commit request) o validación de código (pull request).
3. **Entrenamiento del modelo:** Alimentar el sistema con ejemplos de código seguro e inseguro para mejorar la precisión de la detección.
4. **Análisis y clasificación:** Utilizar la IA para analizar el código, identificar vulnerabilidades potenciales y clasificarlas por severidad.
5. **Retroalimentación a desarrolladores:** Configurar el sistema para proporcionar retroalimentación detallada y sugerencias de corrección a los desarrolladores.

Configuración segura automatizada

Para realizar una configuración segura de sistemas y aplicaciones, reduciendo el riesgo de errores humanos y asegurando el cumplimiento de las mejores prácticas de seguridad.

CASO PRÁCTICO: Desarrollo de un sistema de configuración segura automatizada.

1. **Definición de políticas:** Establecer las políticas de seguridad y las configuraciones deseadas para diferentes tipos de sistemas.
2. **Creación de modelos:** Utilizar aprendizaje automático para crear modelos que puedan evaluar y ajustar configuraciones.
3. **Integración con herramientas de gestión:** Conectar el sistema con herramientas de gestión de configuración y orquestación.
4. **Implementación y monitorización:** Desplegar el sistema en un entorno controlado y monitorizar su rendimiento.
5. **Mejora continua:** Utilizar la retroalimentación de la monitorización para mejorar los modelos y las políticas de configuración.

Evaluación de riesgos de seguridad en nuevos proyectos

Para ayudar en la evaluación de los riesgos de seguridad asociados con nuevos proyectos o tecnologías, permitiendo una toma de decisiones más informada.

CASO PRÁCTICO: Implementación de un sistema de evaluación de riesgos basado en IA.

1. **Recopilación de datos:** Agregar información sobre proyectos anteriores, incidentes de seguridad y mejores prácticas de la industria.
2. **Desarrollo del modelo:** Crear un modelo de aprendizaje automático que pueda evaluar los riesgos basándose en las características del proyecto.
3. **Integración en el proceso:** Incorporar la evaluación de riesgos automatizada en el proceso de aprobación de nuevos proyectos.
4. **Generación de informes:** Configurar el sistema para generar informes detallados de riesgos y recomendaciones.
5. **Actualización continua:** Mantener el modelo actualizado con nueva información y tendencias de seguridad.

EQUIPO AMARILLO (YELLOW TEAM):

CODIFICADORES Y ARQUITECTOS DE SOFTWARE. "LOS CONSTRUCTORES"

El Yellow Team es el equipo que se centra en el diseño, desarrollo e implementación de infraestructuras y sistemas de seguridad robustos. Trabajan

en colaboración con otros equipos de seguridad para construir y mantener una sólida postura de seguridad en la organización.

A continuación, se detalla cómo puede utilizar la IA para aumentar la eficiencia y eficacia de las operaciones de seguridad:

Automatización de flujos de trabajo de seguridad

Para identificar y automatizar procesos repetitivos, liberando tiempo valioso para tareas más estratégicas.

CASO PRÁCTICO: Implementación de un sistema de automatización de flujos de trabajo basado en IA.

1. **Análisis de procesos:** Identificar los procesos de seguridad que son candidatos para la automatización.
2. **Diseño de flujos de trabajo:** Crear flujos de trabajo detallados para cada proceso seleccionado.
3. **Implementación de IA:** Utilizar algoritmos de aprendizaje automático para optimizar los flujos de trabajo y tomar decisiones en puntos clave.
4. **Integración con sistemas existentes:** Conectar el sistema de automatización con las herramientas y plataformas de seguridad existentes.
5. **Monitorización y mejora:** Supervisar el rendimiento de los flujos de trabajo automatizados y ajustar los modelos según sea necesario.

Optimización de recursos de seguridad

Para ayudar a optimizar la asignación de recursos de seguridad, asegurando que se utilicen de la manera más eficiente posible.

CASO PRÁCTICO: Desarrollo de un sistema de optimización de recursos basado en IA.

1. **Recopilación de datos:** Agregar información sobre el uso de recursos, incidentes de seguridad y métricas de rendimiento.
2. **Modelado predictivo:** Crear modelos de aprendizaje automático que puedan predecir las necesidades de recursos basándose en patrones históricos y tendencias actuales.

3. **Implementación de algoritmos de optimización:** Utilizar técnicas como la programación lineal o los algoritmos genéticos para optimizar la asignación de recursos.
4. **Integración con sistemas de gestión:** Conectar el sistema de optimización con las herramientas de gestión de recursos y planificación.
5. **Evaluación y ajuste:** Monitorizar el impacto de las recomendaciones de optimización y ajustar los modelos según sea necesario.

Mejora continua de políticas de seguridad

Para facilitar el análisis de la efectividad de las políticas de seguridad existentes y sugerir mejoras basadas en datos.

CASO PRÁCTICO: Implementación de un sistema de mejora continua de políticas basado en IA.

1. **Recopilación de datos:** Agregar información sobre incidentes de seguridad, cumplimiento de políticas y feedback de usuarios.
2. **Análisis de efectividad:** Utilizar técnicas de aprendizaje automático para evaluar la efectividad de las políticas existentes.
3. **Generación de recomendaciones:** Configurar el sistema para generar recomendaciones de mejora basadas en el análisis.
4. **Simulación de impacto:** Utilizar modelos predictivos para simular el impacto potencial de los cambios de política.
5. **Implementación y seguimiento:** Implementar los cambios recomendados y monitorizar su impacto a lo largo del tiempo.

EQUIPO NARANJA (ORANGE TEAM):

FACILITAN LA INTERACCIÓN Y LA EDUCACIÓN EN CIBERSEGURIDAD

El Orange Team, se enfoca en educar y concienciar a los empleados y usuarios de una organización sobre las mejores prácticas de seguridad cibernética y los riesgos asociados. Su objetivo es promover una cultura de seguridad y educar a través de la formación, para ayudar a prevenir y mitigar las amenazas y los errores humanos que pueden conducir a incidentes de seguridad.

A continuación, se detalla cómo puede aprovechar la IA para promover una cultura de seguridad en las organizaciones:

Formación personalizada basada en IA

Ayuda a analizar el comportamiento y las interacciones de los empleados con los sistemas para crear programas de formación personalizados.

CASO PRÁCTICO: Desarrollo de un sistema de formación personalizada

1. **Recopilación de datos:** Recopilar datos sobre las interacciones de los empleados con los sistemas de la empresa.
2. **Identificación de patrones:** Utilizar algoritmos de machine learning para identificar patrones de comportamiento de riesgo.
3. **Creación de módulos formativos:** Crear módulos de formación específicos que aborden las debilidades detectadas.
4. **Implementación de un sistema de aprendizaje adaptativo:** Implementar un sistema de aprendizaje adaptativo que ajuste el contenido en tiempo real según el progreso del empleado.

Beneficios:

- Formación más relevante y efectiva para cada individuo.
- Mejor retención de conocimientos al abordar áreas específicas de mejora.
- Optimización del tiempo de formación al enfocarse en las necesidades reales de cada empleado.

Simulaciones de ataques basadas en IA

Permite generar escenarios de phishing y otros ataques simulados, adaptados al contexto específico de la organización.

CASO PRÁCTICO: Simulación de un ataque.

1. **Análisis de datos:** Utilizar modelos de IA para analizar los patrones de comunicación interna de la empresa.
2. **Generación del vector de ataque:** Generar correos electrónicos de phishing personalizados que imiten el estilo y contenido habitual de la organización.
3. **Monitorización de la respuesta:** Implementar un sistema de seguimiento que registre las respuestas de los empleados a estas simulaciones.

4. **Recopilación de la información:** Utilizar algoritmos de aprendizaje automático para analizar las respuestas y ajustar futuras simulaciones.

Beneficios:

- Simulaciones más realistas y difíciles de detectar.
- Mejora continua en la capacidad de los empleados para identificar amenazas.
- Datos valiosos sobre las áreas de vulnerabilidad en la organización.

Asistente virtual de seguridad impulsado por IA

Desarrollar un chatbot basado en IA puede proporcionar respuestas instantáneas a preguntas sobre políticas de seguridad y mejores prácticas.

CASO PRÁCTICO: Desarrollo de un chatbot

1. **Generación de una base de conocimiento:** Desarrollar una base de conocimientos que incluya todas las políticas y procedimientos de seguridad de la organización.
2. **Procesamiento de lenguaje natural:** Implementar un modelo de procesamiento de lenguaje natural para interpretar las preguntas de los usuarios.
3. **Aprendizaje continuo:** Utilizar algoritmos de aprendizaje automático para mejorar continuamente la precisión de las respuestas.
4. **Integración con plataformas existentes:** Integrar el asistente en las plataformas de comunicación interna de la empresa.

Beneficios:

- Acceso instantáneo a información crucial de seguridad.
- Reducción de la carga de trabajo para el equipo de soporte técnico.
- Mejora continua en la calidad de las respuestas basada en las interacciones de los usuarios.

EQUIPO ROJO (RED TEAM):

SEGURIDAD OFENSIVA “LOS ATACANTES”

El Red Team, es el responsable de simular ataques reales para identificar vulnerabilidades en los sistemas y las redes de una organización. Trabajan

desde una perspectiva de atacante para mejorar la postura de seguridad y ayudar a fortalecer las defensas.

A continuación, se detalla cómo puede utilizar la IA para crear escenarios de ataque más sofisticados y realistas.

Generación de ataques personalizados

Permite generar ataques personalizados basados en las características específicas de la organización objetivo.

CASO PRÁCTICO: Desarrollo de un sistema de generación de ataques personalizados con IA.

1. **Recopilación de información:** Agregar datos sobre la infraestructura, aplicaciones y políticas de seguridad de la organización objetivo.
2. **Modelado de amenazas:** Utilizar técnicas de aprendizaje automático para modelar posibles vectores de ataque.
3. **Generación de escenarios:** Configurar el sistema para generar escenarios de ataque detallados basados en el modelo de amenazas.
4. **Simulación y refinamiento:** Ejecutar simulaciones de los ataques generados y refinar los modelos basándose en los resultados.
5. **Documentación y reporting:** Generar informes detallados de los ataques simulados y las vulnerabilidades identificadas.

Evasión de defensas basadas en IA

Permite desarrollar métodos que evadan las defensas basadas en IA, probando así la robustez de estas defensas.

CASO PRÁCTICO: Desarrollo de técnicas de evasión de IA utilizando aprendizaje adversario.

1. **Análisis de defensas:** Estudiar las defensas basadas en IA utilizadas por la organización.
2. **Diseño de modelos adversarios:** Crear modelos de aprendizaje automático diseñados para generar entradas que engañen a los sistemas de defensa.
3. **Entrenamiento iterativo:** Entrenar los modelos adversarios contra las defensas, refinando sus técnicas de evasión.

4. **Pruebas de penetración:** Utilizar las técnicas de evasión desarrolladas en pruebas de penetración reales.
5. **Análisis y reporte:** Documentar las técnicas de evasión de éxito y proporcionar recomendaciones para mejorar las defensas.

Automatización de reconocimiento y enumeración

Permite automatizar y optimizar las fases de reconocimiento y enumeración en sus operaciones.

CASO PRÁCTICO: Implementación de un sistema de reconocimiento y enumeración automatizado con IA.

1. **Definición de objetivos:** Establecer los tipos de información y sistemas que se buscan durante el reconocimiento.
2. **Desarrollo de crawlers inteligentes:** Crear crawlers web y de red que utilicen IA para navegar y recopilar información de manera eficiente.
3. **Análisis de datos:** Utilizar técnicas de procesamiento de lenguaje natural y análisis de imágenes para extraer información relevante de los datos recopilados.
4. **Correlación y mapeo:** Emplear algoritmos de aprendizaje automático para correlacionar la información y crear mapas detallados de la infraestructura objetivo.
5. **Generación de informes:** Configurar el sistema para generar informes detallados de los hallazgos del reconocimiento y la enumeración.

EQUIPO PÚRPURA (PURPLE TEAM):

INTEGRAR TÁCTICAS DEFENSIVAS CON RESULTADOS OFENSIVOS

El Purple Team, combina las perspectivas del Blue Team y el Red Team. Se enfoca en realizar ejercicios de simulación de ataques controlados, donde el equipo rojo ataca y el equipo azul defiende. Luego, ambos equipos analizan los resultados y colaboran para mejorar las defensas y la capacidad de respuesta.

A continuación, se detalla cómo puede utilizar la IA para mejorar la colaboración y el intercambio de conocimientos entre estos equipos.

Análisis automatizado de brechas de seguridad

Ayuda a identificar y analizar las brechas de seguridad de manera más eficiente, comparando las tácticas del Red Team con las defensas del Blue Team.

CASO PRÁCTICO: Implementación de un sistema de análisis de brechas basado en IA.

1. **Recopilación de datos:** Agregar información sobre los ataques simulados por el Red Team y las respuestas del Blue Team.
2. **Modelado de brechas:** Utilizar algoritmos de aprendizaje automático para identificar patrones en las brechas de seguridad detectadas.
3. **Análisis predictivo:** Desarrollar modelos que puedan predecir posibles brechas futuras basándose en tendencias históricas.
4. **Generación de recomendaciones:** Configurar el sistema para proporcionar recomendaciones específicas para cerrar las brechas identificadas.
5. **Seguimiento y mejora continua:** Monitorizar la implementación de las recomendaciones y su impacto en la postura de seguridad.

Simulación de escenarios de ataque-defensa

Permite crear y ejecutar simulaciones complejas de escenarios de ataque-defensa, permitiendo un entrenamiento más realista y efectivo.

CASO PRÁCTICO: Desarrollo de un sistema de simulación de ataque-defensa con IA.

1. **Diseño de escenarios:** Crear una biblioteca de escenarios de ataque basados en amenazas reales y emergentes.
2. **Modelado de comportamiento:** Utilizar técnicas de aprendizaje por refuerzo para modelar el comportamiento de atacantes y defensores.
3. **Simulación en tiempo real:** Implementar un entorno de simulación que permita la interacción en tiempo real entre los modelos de ataque y defensa.
4. **Análisis de resultados:** Utilizar algoritmos de IA para analizar los resultados de las simulaciones y extraer insights clave.
5. **Retroalimentación y mejora:** Utilizar los resultados para mejorar tanto las tácticas de ataque como las estrategias de defensa.

Optimización de la transferencia de conocimientos

Facilita la transferencia de conocimientos entre el Red Team y el Blue Team, ayudando a cerrar la brecha entre las perspectivas ofensivas y defensivas.

CASO PRÁCTICO: Implementación de un sistema de gestión del conocimiento basado en IA.

1. **Recopilación de información:** Agregar datos de informes, logs y documentación de ambos equipos.
2. **Procesamiento de lenguaje natural:** Utilizar técnicas de PLN para extraer y categorizar información clave.
3. **Generación de contenido:** Emplear modelos de lenguaje para generar resúmenes y explicaciones de conceptos complejos.
4. **Sistema de recomendación:** Desarrollar un sistema que recomiende recursos relevantes basados en el contexto y las necesidades del usuario.
5. **Aprendizaje continuo:** Actualizar constantemente la base de conocimientos con nueva información y feedback de los usuarios.

CONCLUSIONES

La integración de la inteligencia artificial en la ciberseguridad ha transformado significativamente la manera en que las organizaciones protegen sus activos digitales. Esta tecnología ha permitido a los diferentes equipos de seguridad mejorar sus capacidades y eficiencia en la detección, prevención y respuesta a amenazas cada vez más sofisticadas.

La IA ha demostrado ser una herramienta versátil y poderosa, capaz de analizar grandes volúmenes de datos en tiempo real, automatizar tareas repetitivas y proporcionar conocimiento valioso para la toma de decisiones estratégicas. Desde la detección de amenazas avanzadas hasta la formación personalizada en seguridad, la IA ha encontrado aplicaciones en todos los aspectos de la ciberseguridad.

Sin embargo, es importante reconocer que la IA no es una solución mágica. Su efectividad depende de la calidad de los datos con los que se entrena y de la experiencia de los profesionales que la implementan y gestionan.

Además, a medida que los ciberdelincuentes también adoptan técnicas de IA, la carrera armamentista digital continúa evolucionando.

En el futuro, es probable que veamos una integración aún más profunda de la IA en las estrategias de ciberseguridad, con un enfoque en la mejora continua, la adaptabilidad y la resiliencia frente a las amenazas emergentes.

BIBLIOGRAFÍA

1. Brundage, M., et al. (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. arXiv preprint arXiv:1802.07228.
2. Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*, 18(2), 1153-1176.
3. Dilek, S., Çakır, H., & Aydın, M. (2015). Applications of artificial intelligence techniques to combating cyber crimes: A review. *International Journal of Artificial Intelligence & Applications*, 6(1), 21-39.
4. Kaloudi, N., & Li, J. (2020). The AI-based cyber threat landscape: A survey. *ACM Computing Surveys (CSUR)*, 53(1), 1-34.
5. Li, J. H. (2018). Cyber security meets artificial intelligence: a survey. *Frontiers of Information Technology & Electronic Engineering*, 19(12), 1462-1474.
6. Mittal, S., et al. (2019). A survey of machine and deep learning techniques for cyber security. *SN Computer Science*, 1(1), 1-20.
7. Nguyen, T. T., & Reddi, V. J. (2019). Deep reinforcement learning for cyber security. arXiv preprint arXiv:1906.05799.
8. Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. In *2010 IEEE symposium on security and privacy* (pp. 305-316). IEEE.
9. Taddeo, M., McCutcheon, T., & Floridi, L. (2019). Trusting artificial intelligence in cybersecurity is a double-edged sword. *Nature Machine Intelligence*, 1(12), 557-560.
10. Xin, Y., et al. (2018). Machine learning and deep learning methods for cybersecurity. *IEEE Access*, 6, 35365-35381. 

DESAFÍOS E IMPACTO DE LA IA

Enric Montesa

LA INTELIGENCIA ARTIFICIAL COMO TECNOLOGÍA DE PROPÓSITO GENERAL

Las tecnologías de propósito general (TPG) son innovaciones revolucionarias que generan un impacto transformador y de amplio alcance en múltiples sectores económicos y sociales. Estas tecnologías destacan por su adaptabilidad y potencial para redefinir numerosos aspectos de la actividad humana.

Características clave de las TPG:

1. Amplia aplicabilidad en diversos campos
2. Capacidad de mejora continua
3. Facilitación de innovaciones complementarias

La Inteligencia Artificial (IA) se ha consolidado como una TPG contemporánea, uniéndose a una lista selecta de tecnologías que han moldeado significativamente el curso de la historia humana. Algunas de las TPG más influyentes a lo largo del tiempo incluyen:

1. La rueda (Prehistoria)
2. La escritura (c. 3200 a.C.)
3. La imprenta (c. 1440 d.C.)
4. La máquina de vapor (s. XVIII)
5. La electricidad (s. XIX)
6. El motor de combustión interna (finales del s. XIX)
7. El ordenador (s. XX)
8. Internet (finales del s. XX)
9. La Inteligencia Artificial (s. XXI)

La IA, como TPG emergente, va a revolucionar numerosos sectores y redefinir la forma en que vivimos y trabajamos en el siglo XXI.



Imagen generada con IA (Dall-e)

EL IMPACTO TRANSFORMADOR DE LA INTELIGENCIA ARTIFICIAL: PERSPECTIVAS DE KAI-FU LEE

Kai-Fu Lee, reconocido experto en inteligencia artificial (IA), ofrece en su libro "Superpotencias de la inteligencia artificial"¹ una visión clara y profunda sobre los cambios fundamentales que la IA está introduciendo y continuará generando en nuestra sociedad. Lee identifica cinco áreas principales de transformación:

1. Disrupción en el Mercado Laboral

- **Automatización:** La IA reemplazará tareas repetitivas y basadas en datos.
- **Reestructuración:** Desaparición de ciertos empleos y creación de otros nuevos.
- **Resistencia:** Trabajos que requieren creatividad, pensamiento crítico y empatía serán menos susceptibles a la automatización.

2. Incremento de la Desigualdad

- **Concentración de beneficios:** Grandes empresas y gobiernos, principales desarrolladores de IA, obtendrán ventajas significativas.

1 Kai-Fu Lee (2020): *Superpotencias de la inteligencia artificial: China, Silicon Valley y el nuevo orden mundial*. Barcelona. Deusto.

- **Brecha tecnológica:** Riesgo de marginación para quienes no tengan acceso a la tecnología o a los datos necesarios.

3. Evolución en las Interacciones Sociales

- **Nuevas herramientas:** Asistentes virtuales, chatbots y otras aplicaciones de IA se integran en la vida cotidiana.
- **Personalización:** Mayor adaptación de información y servicios a las necesidades individuales.
- **Cambios sociales:** Posible surgimiento de nuevas formas de relaciones interpersonales mediadas por la IA.

4. Desafíos Éticos

- **Cuestiones clave:** Privacidad, seguridad, responsabilidad y equidad en el uso de la IA.
- **Regulación:** Necesidad de establecer marcos éticos y regulatorios sólidos.
- **Objetivo:** Garantizar un desarrollo responsable y beneficioso de la IA para toda la sociedad.

5. Competencia Geopolítica

- **Liderazgo global:** Estados Unidos y China compiten por la supremacía en IA.
- **Impacto global:** Esta competencia influirá significativamente en la economía mundial.
- **Equilibrio de poder:** Posibles cambios en las dinámicas geopolíticas internacionales.

Estas transformaciones representan tanto oportunidades como desafíos para la sociedad global. Es crucial fomentar un diálogo abierto y una colaboración internacional para maximizar los beneficios de la IA y mitigar sus potenciales riesgos.

ADOPCIÓN DE LA INTELIGENCIA ARTIFICIAL POR LAS ORGANIZACIONES

La integración de la Inteligencia Artificial (IA) en las organizaciones está transformando de forma radical la forma en que operan, toman decisiones y crean valor. Esta adopción está impulsando una metamorfosis digital sin precedentes en diversos sectores.

Objetivos Principales de la Adopción de IA

1. Automatización de procesos
2. Reducción de costos operativos
3. Mejora de la eficiencia operativa
4. Optimización de la toma de decisiones

Tecnologías Clave y sus Aplicaciones

Machine Learning

- Predicción de fallos en equipos
- Optimización de la cadena de suministro
- Personalización de productos y servicios

Analítica Avanzada

- Análisis de tendencias de mercado
- Segmentación de clientes
- Detección de fraudes

Procesamiento de Lenguaje Natural (PLN)

- Chatbots para atención al cliente
- Análisis de sentimientos en redes sociales
- Automatización de procesos de documentación

Beneficios de la Implementación de IA

1. **Análisis de Big Data:** Capacidad para procesar y analizar grandes volúmenes de datos en tiempo real.
2. **Toma de Decisiones Ágil:** Facilita decisiones informadas y rápidas basadas en datos actualizados.
3. **Reducción de Riesgos:** Mejora la previsión y gestión de riesgos mediante análisis predictivos.
4. **Mejora de la Experiencia del Cliente:** Personalización y atención al cliente mejoradas.
5. **Innovación:** Facilita el desarrollo de nuevos productos y servicios basados en insights derivados de datos.

Desafíos en la Adopción de IA

- **Inversión Inicial:** Costos asociados con la implementación de tecnologías de IA.
- **Resistencia al Cambio:** Necesidad de gestionar la transición y capacitar al personal.
- **Ética y Privacidad:** Garantizar un uso ético de los datos y proteger la privacidad de los usuarios.
- **Integración con Sistemas Existentes:** Asegurar la compatibilidad con la infraestructura tecnológica actual.

En definitiva, la adopción de la IA por parte de las organizaciones representa un cambio paradigmático en la forma de operar y competir. Aquellas que logren integrar eficazmente estas tecnologías estarán mejor posicionadas para enfrentar los desafíos del mercado actual y futuro, mejorando su competitividad y capacidad de innovación.

LA INTELIGENCIA ARTIFICIAL COMO CATALIZADOR DE “NUEVOS MODELOS DE NEGOCIO”

La Inteligencia Artificial (IA) está transformando radicalmente el panorama empresarial, impulsando el crecimiento de modelos de negocio innovadores y redefiniendo las estrategias de monetización de datos. Este cambio paradigmático está reconfigurando industrias enteras y creando nuevos ecosistemas empresariales.

Modelos de Negocio Impulsados por IA

1. *Monetización de Datos*

La IA ha potenciado la capacidad de las empresas para extraer valor de los datos, dando lugar a modelos de negocio centrados en:

- **Publicidad Personalizada:** Algoritmos que analizan el comportamiento del usuario para ofrecer anuncios altamente dirigidos.
- **Sistemas de Recomendación:** Plataformas que sugieren productos o contenidos basados en preferencias individuales.

- **Gestión Dinámica de Precios:** Ajuste en tiempo real de precios basado en la demanda, la oferta y otros factores del mercado.

2. *Interacción Mejorada con el Cliente*

La IA está revolucionando la forma en que las empresas se comunican con sus clientes a través de:

- **Asistentes Virtuales:** Capaces de manejar consultas complejas y proporcionar asistencia personalizada.
- **Interfaces Conversacionales:** Chatbots y sistemas de procesamiento del lenguaje natural que facilitan interacciones más naturales.
- **Plataformas de Atención al Cliente Automatizadas:** Sistemas que manejan grandes volúmenes de consultas, mejorando la eficiencia y la satisfacción del cliente.

El Auge de las Plataformas Basadas en IA

Empresas tecnológicas líderes como Amazon, Google y Alibaba han aprovechado la IA para crear poderosas plataformas que:

- **Crean Ecosistemas Empresariales:** Facilitan la colaboración y la competencia entre diversas empresas en torno a tecnologías emergentes.
- **Fomentan la Innovación:** Proporcionan herramientas y recursos para que otras empresas desarrollen soluciones basadas en IA.
- **Dominan Múltiples Sectores:** Expanden su influencia más allá de sus áreas originales de negocio.

Este fenómeno ha llevado a algunos expertos, como Yanis Varoufakis en su libro "*Tecnofeudalismo: El sigiloso sucesor del capitalismo*"², a argumentar que estamos presenciando la aparición de un nuevo orden económico. Según esta perspectiva, las grandes plataformas tecnológicas están acumulando un poder sin precedentes, comparable al de los señores feudales de la Edad Media, pero en el ámbito digital.

2 Varoufakis, Yanis (2024). *Tecnofeudalismo: El sigiloso sucesor del capitalismo*. Bilbao: Deusto

Implicaciones y Desafíos

- **Concentración de Poder:** La dominancia de las grandes plataformas de IA plantea preocupaciones sobre la competencia y la innovación a largo plazo.
- **Privacidad y Ética:** El uso intensivo de datos personales para alimentar estos modelos de negocio suscita debates éticos y regulatorios.
- **Transformación del Empleo:** La automatización impulsada por IA está redefiniendo los roles laborales y las habilidades requeridas en el mercado laboral.
- **Adaptación Empresarial:** Las empresas tradicionales se ven obligadas a transformar sus modelos de negocio para mantenerse competitivas en la era de la IA.

Las conclusiones son evidentes. La IA no solo está mejorando los modelos de negocio existentes, sino que está creando otros completamente nuevos. Se trata de un cambio profundo que ofrece oportunidades sin precedentes, pero que también plantea desafíos importantes para las empresas, los reguladores y la sociedad en general. La capacidad de adaptarse a este nuevo paradigma será crucial para el éxito en la economía digital del futuro.

IMPACTO DE LA INTELIGENCIA ARTIFICIAL SOBRE EL EMPLEO Y LAS PROFESIONES

La Inteligencia Artificial (IA) está transformando radicalmente el panorama laboral, generando tanto oportunidades como desafíos para trabajadores y empresas. Este cambio tecnológico sin precedentes ha sido objeto de extenso análisis por parte de expertos en economía, tecnología y política pública.

Automatización y Desplazamiento Laboral

Uno de los efectos más discutidos de la IA es su capacidad para automatizar tareas tradicionalmente realizadas por humanos. Jerry Kaplan, en su libro *"Abstenerse humanos: Guía para la riqueza y el trabajo en la era de la inteligencia artificial"* (2016), argumenta que la IA y la robótica están creando una nueva clase de "trabajadores digitales" que pueden realizar una amplia

gama de tareas cognitivas y físicas, potencialmente desplazando a millones de trabajadores humanos³.

Sin embargo, Erik Brynjolfsson y Andrew McAfee, en *"La segunda era de las máquinas: Trabajo, progreso y prosperidad en una época de brillantes tecnologías"* (2014), ofrecen una perspectiva más matizada. Sostienen que si bien la IA ciertamente automatizará muchos trabajos, también creará nuevas oportunidades y aumentará la productividad en formas que podrían beneficiar a la economía en general⁴.

Transformación de Profesiones

La IA no solo está eliminando trabajos, sino que está transformando profundamente muchas profesiones. David Autor, economista del MIT, ha estudiado cómo la tecnología complementa ciertas habilidades humanas mientras sustituye otras⁵. En su trabajo *"Why Are There Still So Many Jobs?"* (2015)⁶, David Autor argumenta que la IA tenderá a aumentar la demanda de trabajos que requieren creatividad, resolución de problemas y habilidades interpersonales complejas.

El trabajo en los próximos diez años

La información aparecida en los medios de comunicación puede sintetizarse en forma de tabla que en la que se muestran algunas de las tareas que la IA probablemente automatizará en los próximos años, así como una estimación del porcentaje de tareas que podrían desaparecer en los próximos 10 años:

3 Kaplan, J. (2016): *Abstenerse humanos: Guía para la riqueza y el trabajo en la era de la inteligencia artificial*. Barcelona: Teell Editorial.

4 Brynjolfsson, E., & McAfee, A. (2014): *La segunda era de las máquinas: Trabajo, progreso y prosperidad en una época de brillantes tecnologías*. Buenos Aires. Temas Grupo Editorial.

5 Autor, D., Mindell, D. A., & Reynolds, E. B. (2020). *The Work of the Future: Building Better Jobs in an Age of Intelligent Machines*. Cambridge, MA: The MIT Press

6 Autor, D. H. (2015). *Why Are There Still So Many Jobs? The History and Future of Workplace Automation*. Journal of Economic Perspectives, 29(3), 3-30 https://www.researchgate.net/publication/282320407_Why_Are_There_Still_So_Many_Jobs_The_History_and_Future_of_Workplace_Automation

ÁREA	PORCENTAJE ESTIMADO DE AUTOMATIZACIÓN
Tareas administrativas y de oficina	30-45% para 2030-2035
Pago a proveedores	46% de grandes empresas planean automatizar
Elaboración de facturas	46% de grandes empresas planean automatizar
Presentación de informes financieros	46% de grandes empresas planean automatizar
Tareas creativas (redacción, MK)	No especificado, pero en aumento
Empleos minoristas	65% para 2025
Tareas manuales y de conducción	Mayor impacto a largo plazo
Producción	Disminución esperada hasta 2030
Servicio al cliente	Disminución esperada hasta 2030
Ventas	Disminución esperada hasta 2030

IMPACTO EN LOS SECTORES ECONÓMICOS CLAVE

💬 José Ortega

Son muchos los sectores que se están beneficiando por la adopción operativa de la IA, y se espera que los próximos años sus aplicaciones se expandan y evolucionen aún más, ofreciendo nuevas mejoras y oportunidades.

Se calcula que el PIB de aquellos países que utilizan la IA de manera generalizada puede incrementarse, como consecuencia de ello, en más de un 7% en pocos años. Se trata de un crecimiento acumulativo que puede crear diferencias insalvables entre países, según su comportamiento ante la IA.

A continuación mencionamos los sectores emergentes y tradicionales que están adoptando esta tecnología así como aquellos que están obteniendo nuevos avances en IA.

AGRICULTURA

Monitoreo de Cultivos y Detección de Enfermedades

- **Visión por Computadora y Drones:** Los drones pueden escanear grandes áreas de cultivo, capturando imágenes que se analizan con algoritmos de IA para detectar problemas como enfermedades, plagas y deficiencias nutricionales. Empresas que lo utilizan: **Taranis** y **SkySquirrel Technologies**.
- **Modelos Predictivos:** Analizan patrones históricos y condiciones climáticas para predecir brotes de enfermedades y plagas, permitiendo a los agricultores tomar medidas preventivas.

Agricultura de Precisión

- **Sensores y Big Data:** Sensores en los campos recopilan datos sobre el suelo, el clima y la salud de los cultivos y se combinan con técnicas de análisis de Big Data para optimizar el riego, la fertilización y otros insumos agrícolas. Empresas: **John Deere**.
- **Robótica y Automatización:** Robots equipados con IA pueden realizar tareas como la siembra, el deshierbe y la cosecha con una precisión y eficiencia superiores a las humanas. Blue River Technology, por ejemplo, ha

desarrollado robots que utilizan IA para identificar y eliminar malezas de manera selectiva.

Optimización de Cosechas

- **Modelos basados en IA** pueden predecir el rendimiento de los cultivos en función de diferentes variables ayudando a tomar decisiones informadas sobre la gestión de los campos.
- **Cosecha Automatizada:** Sistemas automatizados de recolección utilizan IA para determinar el momento óptimo de cosecha. Empresas: **Harvest CROO Robotics**.

Gestión del Agua y Recursos

- **Riego Inteligente:** Predicciones climáticas basadas en IA se utilizan para optimizar el uso del agua. Empresas: **Netafim** y **CropX**.
- **Gestión de Recursos:** Para la gestión eficiente de otros recursos como fertilizantes y pesticidas, y el uso de cantidades adecuadas en el momento correcto, reduciendo costes e impacto ambiental.

Plataformas de Gestión Agrícola

- **Software de Gestión:** Herramientas para gestionar y analizar todos los aspectos de las operaciones agrícolas. Empresas: **FarmLogs** y **Climate FieldView**.
- **Asistentes Virtuales:** Para ayudar a los agricultores a tomar decisiones informadas proporcionándoles recomendaciones personalizadas y en tiempo real sobre la gestión de sus cultivos.

ENERGÍA

Optimización de Redes Eléctricas

- **Redes Inteligentes (Smart Grids):** Para optimizar la oferta y la demanda de electricidad en tiempo real.
- **Detección y Respuesta a Fallos:** Para ayudar a detectar y diagnosticar fallos en la red eléctrica de manera rápida y precisa.

Gestión de la Demanda Energética

- **Predicción de la Demanda:** Para anticipar los patrones de consumo de energía, permitiendo ajustar la producción y distribución de electricidad de manera eficiente.
- **Programas de Respuesta a la Demanda:** Para que consumidores y empresas participen en programas de respuesta a la demanda y ajustar el consumo en función de las señales del mercado y las condiciones de la red.

Energías Renovables

- **Optimización de la Producción:** Se utiliza para optimizar la producción. Algoritmos avanzados analizan datos meteorológicos y operativos para predecir la generación de energía y ajustar las operaciones en consecuencia.
- **Mantenimiento Predictivo:** La IA permite el mantenimiento predictivo de las instalaciones de energías renovables, identificando posibles fallos antes de que ocurran.

Almacenamiento de Energía

- **Optimización del Uso de Baterías:** Para optimizar los sistemas de almacenamiento de energía, como baterías, optimizando su carga y descarga en función de las condiciones de la red y los precios de la energía.

Eficiencia Energética en Edificios

- **Gestión de Energía en Edificios:** La IA se utiliza en sistemas de gestión de energía para edificios, optimizando el consumo de electricidad, calefacción y refrigeración en función de las condiciones ambientales y las necesidades de los ocupantes. Empresas: como **Siemens** y **Honeywell**.
- **Automatización de Sistemas:** Para la automatización de sistemas de iluminación, climatización y otros dispositivos en edificios, ajustándolos automáticamente para reducir el consumo de energía sin comprometer el confort.

Análisis de Datos y Toma de Decisiones

- **Plataformas de Análisis de Energía:** Herramientas de análisis como Google's DeepMind y IBM's Watson proporcionan a las empresas de energía insights detallados sobre sus operaciones, ayudándoles a tomar decisiones.
- **Predicción de Precios de Energía:** Se utiliza para predecir los precios de la energía en los mercados.

VEHÍCULOS AUTÓNOMOS

Mejora en la Percepción y Sensores

- **Fusión de Sensores:** Algoritmos avanzados de IA permiten una fusión de sensores más precisa, mejorando la detección de obstáculos, peatones y otros vehículos.
- **Detección y Clasificación Mejoradas:** Los algoritmos de visión por computadora y aprendizaje profundo han mejorado significativamente la capacidad de los vehículos autónomos para detectar y clasificar objetos en tiempo real, incluso en condiciones adversas como lluvia, niebla o baja iluminación.

Navegación y Planificación de Rutas

- **Mapeo HD en Tiempo Real:** Los vehículos autónomos ahora pueden crear y actualizar mapas de alta definición en tiempo real.
- **Planificación de Rutas Dinámicas:** Permite a los vehículos autónomos planificar rutas de manera dinámica, ajustándose en tiempo real a cambios en el tráfico, condiciones del camino y eventos imprevistos, optimizando el tiempo de viaje y la eficiencia.

Conducción Segura y Eficiente

- **Conducción Predictiva:** La IA permite a los vehículos autónomos anticipar las acciones de otros usuarios de la vía, como peatones, ciclistas y otros vehículos, mejorando la seguridad y reduciendo la probabilidad de accidentes.
- **Optimización de Consumo de Energía:** Algoritmos de aprendizaje automático optimizan el consumo de energía de vehículos eléctricos autónomos, gestionando de manera eficiente la aceleración, frenado y uso de la batería para maximizar la autonomía.

Interacción Vehículo-Infraestructura

- **Comunicación V2X (Vehicle-to-Everything):** La IA mejora la comunicación entre vehículos autónomos y la infraestructura vial.
- **Infraestructura Inteligente:** Las ciudades están implementando infraestructura inteligente que se comunica con los vehículos autónomos para proporcionar información en tiempo real.

Pruebas y Validación

- **Simulaciones Realistas:** Se ha mejorado las capacidades de simulación, permitiendo pruebas exhaustivas en entornos virtuales que replican condiciones del mundo real.
- **Aprendizaje por Refuerzo:** Los vehículos autónomos utilizan técnicas de aprendizaje por refuerzo para mejorar su comportamiento a través de la experiencia.

INDUSTRIA 4.0

Mantenimiento Predictivo

- **Implementación a Gran Escala:** Empresas como **GE** y **Siemens** han lanzado plataformas que utilizan datos en tiempo real para predecir fallos en maquinaria industrial, reduciendo tiempos de inactividad y costos de mantenimiento.

Gemelos Digitales

- **Avances en Realismo y Precisión:** Los gemelos digitales han mejorado significativamente en realismo y precisión. Ahora se utilizan en la simulación y optimización de líneas de producción completas, permitiendo ajustes y mejoras sin interrumpir la producción real. Empresa: **Siemens** y **PTC**.

Automatización Inteligente

- **Robots Autónomos y Cobots:** La colaboración entre humanos y robots (cobots) ha avanzado, con robots que ahora pueden aprender nuevas tareas mediante técnicas de aprendizaje automático. Empresa: **Universal Robots**.

- **Nuevas Aplicaciones de Robots Autónomos:** Los robots autónomos, equipados con IA avanzada, se están utilizando para tareas más complejas como la inspección de infraestructuras y la logística interna en fábricas.

Inspección Automatizada y Control de Calidad

- **Visión por Computadora Avanzada:** Los sistemas de visión por computadora han mejorado notablemente en precisión y velocidad. Empresa: **Cognex** y **Keyence**.
- **Integración con Líneas de Producción:** La IA se ha integrado más profundamente en las líneas de producción, permitiendo ajustes automáticos en los procesos para mantener la calidad sin intervención humana constante.

Optimización de la Cadena de Suministro

- **Algoritmos Predictivos para Gestión de Inventarios:** Como predecir la demanda y gestionar inventarios de manera más eficiente. Empresas: **IBM** y **SAP**.
- **Resiliencia en la Cadena de Suministro:** La pandemia de COVID-19 impulsó el uso de la IA para mejorar la resiliencia de las cadenas de suministro, permitiendo a las empresas adaptarse rápidamente a interrupciones y cambios en la demanda.

Sostenibilidad y Eficiencia Energética

- **Optimización del Uso de Recursos:** Para optimizar el uso de recursos naturales y reducir el consumo de energía en procesos industriales. Plataformas: **OSISoft** y **Uptake**.
- **Iniciativas de Producción Sostenible:** Las empresas han adoptado IA para implementar prácticas de producción más sostenibles, incluyendo la reducción de emisiones y el reciclaje eficiente de materiales.

Seguridad y Salud Ocupacional

- **Monitoreo en Tiempo Real:** Para monitorear el cumplimiento de normas de seguridad en tiempo real, identificando y mitigando riesgos de manera proactiva.
- **Prevención de Accidentes Laborales:** Algoritmos de IA analizan datos de salud y seguridad para predecir y prevenir accidentes laborales.

Personalización y Flexibilidad en la Producción

- **Fabricación Bajo Demanda:** La IA ha permitido una mayor flexibilidad en la producción, facilitando la fabricación bajo demanda y la personalización de productos a gran escala. Empresas: **HP y GE.**
- **Configuración Automática de Máquinas:** Permite la configuración automática de máquinas para diferentes tareas.

TURISMO

Ventajas de la IA en el Turismo

Personalización de la Experiencia del Cliente

- **Recomendaciones Personalizadas:** Los sistemas de IA pueden analizar preferencias y comportamientos pasados de los viajeros para ofrecer recomendaciones personalizadas de destinos, alojamientos, actividades y restaurantes. Empresa: Booking.com, TripAdvisor.
- **Itinerarios a Medida:** Aplicaciones de IA pueden crear itinerarios de viaje personalizados que optimizan el tiempo y los intereses del turista, mejorando su experiencia global. Esto incluye la comparación de precios, la gestión de reservas y la sincronización de horarios de vuelos, alojamiento y actividades. Empresa: Booking.com, Airbnb.
- **Predicción de Disponibilidad y Tarifas:** Los algoritmos de IA pueden predecir la disponibilidad de vuelos, hoteles y actividades, así como las fluctuaciones en las tarifas, permitiendo a los viajeros reservar en el momento más oportuno y económico. Empresa: Kayak.
- **Optimizadores de Rutas:** Herramientas de IA ayudan a planificar las rutas de viaje más eficientes, considerando factores como el tráfico, el tiempo y las condiciones climáticas, optimizando así el tiempo y la experiencia del viaje. Empresa: Expedia.
- **Experiencias en Destinos Más Enriquecedoras. Guías Locales Inteligentes:** Para hacer recomendaciones en tiempo real sobre atracciones locales, restaurantes y eventos, adaptadas a los intereses del viajero. Empresa: TripAdvisor.

Experiencias de Realidad Virtual (VR) y Realidad Aumentada (AR)

- **Tours Virtuales:** Las tecnologías VR y AR permiten a los turistas explorar destinos, museos y sitios históricos virtualmente antes de viajar. Esto no solo mejora la planificación y la anticipación del viaje, sino que también ofrece una alternativa para aquellos que no pueden viajar físicamente.
- **Guías de Turismo Aumentadas:** Durante el viaje, los dispositivos móviles equipados con AR pueden superponer información digital en el entorno real, proporcionando a los turistas datos históricos, culturales y prácticos en tiempo real mientras exploran un lugar.
- **Simulación de Experiencias:** La VR puede simular experiencias turísticas inmersivas, como vuelos en ala delta, buceo en arrecifes de coral o caminatas en sitios remotos, ofreciendo a los turistas una prueba virtual antes de experimentar la actividad real.

Mejora en los servicios a los clientes

- **Mejora en la Atención al Cliente:** Mediante Asistentes Virtuales y Chatbots se puede proporcionar información y asistencia a los clientes en tiempo real.
- **Servicio Multilingüe:** Herramientas de traducción de IA facilitan la comunicación entre turistas y locales o servicios, superando barreras lingüísticas.
- **Resolución Proactiva de Problemas:** La IA puede anticipar y solucionar problemas de los clientes antes de que estos ocurran, como la reprogramación de reservas en caso de cambios de vuelo o recomendaciones de actividades alternativas en caso de mal tiempo.
- **Recomendaciones Ecológicas:** La IA puede promover opciones de viaje más sostenibles al recomendar alojamientos ecológicos, medios de transporte con menor huella de carbono y actividades respetuosas con el medio ambiente.
- **Optimización de Recursos:** En el sector hotelero, la IA ayuda a gestionar de manera más eficiente los recursos energéticos y de agua, reduciendo el impacto ambiental y mejorando la sostenibilidad operativa.
- **Asistencia Sanitaria y Emergencias:** Aplicaciones de IA pueden guiar a los turistas en casos de emergencias médicas, proporcionando información

sobre los centros de salud más cercanos y procedimientos de primeros auxilios.

Optimización de Operaciones y Gestión

- **Mantenimiento Predictivo:** En el sector hotelero y de transporte, la IA puede predecir necesidades de mantenimiento y prevenir fallos, mejorando la fiabilidad y seguridad de los servicios.
- **Segmentación de Mercado:** La IA permite una segmentación de mercado más precisa y campañas de marketing dirigidas, que se adaptan a los intereses y comportamientos específicos de los viajeros.
- **Análisis de Sentimiento:** Las herramientas de IA pueden analizar reseñas y comentarios en línea para evaluar la satisfacción del cliente y ajustar los servicios o la oferta en consecuencia.

Inconvenientes y Desafíos de la IA en el Turismo

- **Recolección de Datos Personales:** La recopilación de datos personales de los viajeros para personalizar experiencias y servicios plantea preocupaciones sobre la privacidad y la seguridad de la información.
- **Uso Ético de Datos:** Existe el riesgo de uso indebido de los datos recopilados, ya sea por violación de la privacidad o por prácticas discriminatorias basadas en perfiles de usuarios.
- **Reducción de Interacción Humana:** El aumento del uso de IA y automatización puede reducir la interacción humana, lo cual es un componente importante de la experiencia turística y de hospitalidad.
- **Fallos Tecnológicos:** La dependencia de la tecnología implica que fallos en los sistemas de IA o interrupciones en el servicio puedan afectar significativamente la experiencia del cliente.
- **Sustitución de Empleos:** La automatización de tareas como la atención al cliente y la gestión de reservas puede llevar a la reducción de empleos.
- **Sesgos en los Algoritmos:** Los sistemas de IA pueden perpetuar sesgos si están entrenados en datos no representativos o sesgados, lo que podría resultar en recomendaciones o servicios que no sean inclusivos o justos.

- **Desigualdad en el Acceso:** No todos los viajeros tienen el mismo acceso a la tecnología necesaria para beneficiarse de las herramientas basadas en IA, lo que puede exacerbar las desigualdades en la experiencia de viaje.

SEGURIDAD Y DEFENSA

Drones y Vehículos Autónomos

- **Drones Inteligentes:** Pueden realizar misiones de reconocimiento, vigilancia y ataque con una mayor precisión y menor intervención humana.
- **Vehículos Terrestres Autónomos:** Pueden operar en entornos hostiles, desactivar explosivos y dar apoyo logístico en el campo de batalla.

Ciberseguridad

- **Detección de Amenazas en Tiempo Real:** Algoritmos de aprendizaje automático analizan patrones de tráfico de red y comportamientos anómalos para identificar posibles ataques antes de que se produzcan. Empresa: **Darktrace**.
- **Análisis de Vulnerabilidades:** Para identificar vulnerabilidades en sistemas y redes.

Reconocimiento de Imágenes y Vigilancia

- **Reconocimiento Facial y de Imágenes:** Los algoritmos de visión por computadora han mejorado en precisión y velocidad, permitiendo la identificación rápida y precisa de individuos en tiempo real. Esto es útil para el control de fronteras, la vigilancia en áreas públicas y la identificación de sospechosos en operaciones de seguridad.
- **Sistemas de Vigilancia Automatizados:** Detectando comportamientos sospechosos y alertando a los operadores humanos para que tomen medidas.

Inteligencia y Análisis de Datos

- **Análisis Predictivo:** La IA ayuda a predecir actividades criminales y amenazas terroristas mediante el análisis de grandes conjuntos de datos provenientes de diversas fuentes, como redes sociales, transacciones financieras y comunicaciones.

- **Integración de Fuentes de Información:** Facilita la integración y el análisis de datos de múltiples fuentes, proporcionando una visión más completa y precisa de las amenazas.

Defensa y Operaciones Militares

- **Sistemas de Defensa Autónomos:** Desarrolla sistemas de defensa autónomos, como misiles y sistemas antiaéreos, que pueden tomar decisiones en fracciones de segundo para interceptar amenazas.
- **Simulaciones y Entrenamiento:** Se utilizan para el entrenamiento de tropas y la planificación de operaciones militares. Estos simuladores pueden recrear escenarios complejos y dinámicos, proporcionando a los soldados una experiencia de entrenamiento realista y adaptable.

Comunicación y Guerra Electrónica

- **Interferencia y Protección de Señales:** Se utiliza para mejorar las capacidades de guerra electrónica y la protección de las propias comunicaciones. Algoritmos avanzados pueden detectar y neutralizar señales hostiles de manera más eficaz.
- **Optimización de Redes de Comunicación:** Asegurando que las transmisiones sean seguras, eficientes y resistentes a las interferencias y ataques cibernéticos.

Robótica y Automatización

- **Robots de Desminado:** Para detectar y desactivar minas terrestres y explosivos improvisados (IEDs).
- **Sistemas de Logística Automatizados:** Para automatizar y optimizar la logística militar, incluyendo el transporte de suministros y la gestión de inventarios en el campo de batalla. 📄

LA IA EN LA COMUNICACIÓN: ADOPTAR LO MEJOR DE LA IA Y GESTIONAR SUS RIESGOS

💬 Mar Monsoriu Flor

En la última década algunos de los grandes medios de comunicación del mundo han estado empleando la Inteligencia Artificial (en adelante IA) para agilizar la producción y las tareas rutinarias: desde la generación de informes y resúmenes deportivos hasta la producción de etiquetas y transcripciones. La cadena de televisión CNN, por ejemplo, lleva tiempo empleando la IA para transcribir automáticamente entrevistas y de esta forma poder publicarlas casi inmediatamente en su web.

A partir de lanzamiento público de ChatGPT el 30 de noviembre de 2022 en la práctica totalidad de las redacciones han comenzado a emplear la IA y, en la actualidad, se está utilizando para llevar a cabo diversos procesos avanzados relacionados con el periodismo como los que se exponen a continuación.

Con la ayuda de la tecnología basada en la IA **las empresas y organizaciones de medios de comunicación pueden digitalizar de manera eficiente sus archivos** y utilizar estos datos como base para futuros informes o investigaciones. Este fácil acceso a los archivos es muy importante particularmente en los medios locales, que abordan noticias que con frecuencia no reciben cobertura nacional y que hasta ahora no han podido contar hasta el momento con una buena hemeroteca.

Además, **en la mayoría de las redacciones se utiliza la IA en la redacción y edición de titulares**. La finalidad es adecuar dichos titulares a la extensión que se desee en cada caso y además hacerlos más atractivos para audiencia. También se persigue la optimización para mejorar el posicionamiento del contenido en los buscadores e incluso la generación automática de breves publicaciones e incluso imágenes y vídeos destinados a las redes sociales.

Por otra parte, **la IA se está empleando para mejorar la relevancia y comprensión de las publicaciones**. Es un paso más allá de la simple corrección lingüística o gramatical como hasta ahora ha venido haciendo el Washington Post. Este medio emplea una herramienta basada en la IA llamada Grammarly (<https://>

www.grammarly.com/) para ayudar a sus editores a revisar los artículos de noticias. Dicha herramienta –disponible sólo en inglés– puede identificar errores gramaticales y ortográficos, así como sugerir formas de mejorar la claridad y la concisión de la escritura.



Otros medios están usando aplicaciones capaces de leer de noticias, resumirlas y reescribirlas para que puedan ser atractivas a diferentes públicos objetivos: desde un niño, hasta un adolescente, pasando por personas con dificultades de comprensión lectora que necesiten apoyarse en muchos dibujos o emojis. Estos asistentes de inteligencia artificial sugieren a los periodistas formas de hacer que las historias sean más interesantes, precisas o relevantes.

Por ejemplo, el 'Hennibot' de Helsingin Sanomat en Finlandia está integrado en su sistema de gestión de contenido (CMS) e indica a los periodistas qué pueden mejorar tanto en la redacción como en los enlaces externos. Según Esa Mäkinen, editor jefe de la publicación: "La fiabilidad y la experiencia de usuario de la plataforma de noticias nos han ayudado en Helsingin Sanomat a prestar un mejor servicio a nuestros lectores, por ello hemos logrado crecer en términos de suscripciones, lo que nos da la oportunidad de ofrecer periodismo de alta calidad para una audiencia aún mayor".

En la publicación sueca Aftonbladet un asistente basado en la IA crea líneas de tiempo, cuadros de datos y preguntas y respuestas escritas en el tono de la marca, pero con un resumen para garantizar que "un joven de 18 años" pueda entender fácilmente el contenido. Mientras tanto, en el otro lado del globo, Filipinas, Rappler también aplica la inteligencia artificial al periodismo mediante un conjunto de herramientas llamadas TLDR (derivado de la expresión en línea "demasiado largo; no lo leí") que convierte las historias de la publicación en resúmenes, gráficos y videos.

Junto a la adecuación del contenido a la audiencia, en Alemania, el tabloide Express.de está llevando a cabo la **generación automática de artículos**. Según Thomas Schultz-Homberg, director ejecutivo de DuMont Mediengruppe, el grupo de medios germano que posee varios periódicos y publicaciones incluido el mencionado tabloide: "el uso de la IA en Express.de en su versión digital está suponiendo un aumento significativo de la tasa de clics".

Ese periódico ha creado una “periodista virtual” llamada Klara Indernach. En realidad, bajo ese nombre se publican los textos que crean con la ayuda de la IA. Klara “escribe” más del 5% de las historias publicadas sobre una amplia gama de temas. Los editores humanos aún deciden qué temáticas se cubren y revisan cada pieza de contenido, pero la redacción, la estructura y el tono se externalizan de manera efectiva con ayuda de la IA.



SOBRE LA PERSONA

KLARA IDERNACH

Klara Indernach es el nombre de los textos que creamos con la ayuda de la inteligencia artificial. Si los artículos se generaron en gran medida con la ayuda de la IA, los marcamos en consecuencia. Antes de su publicación, son editados y revisados por el equipo editorial. La foto de perfil se creó con la ayuda de Midjourney.

ULTIMOS ARTICULOS:

Junto a la redacción y edición de los contenidos, la IA está siendo de gran ayuda en la traducción de los mismos como en el caso de Le Monde, en Francia. Este medio utiliza la IA para ayudar a traducir sus artículos y noticias. Eso permite que aparezcan alrededor de 30 artículos al día en su edición en inglés, lo que es mucho más de lo que se estaba haciendo con anterioridad. La IA ayuda con la traducción inicial a la que le siguen varias comprobaciones humanas. El software que se emplea ha sido personalizado para reconocer el libro de estilo y la ortografía de Le Monde.

De forma paralela destaca la transcripción de audio al propio idioma o a otros. Por ejemplo, VerdensGang o VG en Noruega, ha lanzado recientemente su servicio Jojo que transcribe y traduce automáticamente los archivos de audio o vídeo de más de 100 idiomas con Jojo con tecnología de OpenAI Whisper.

Como se ha visto con el ejemplo alemán, **la IA se emplea para generar automáticamente artículos de noticias, resúmenes y otros tipos de contenido.** Esto deja más tiempo a los periodistas que de esta forma pueden centrarse en trabajos de investigación que aporten más valor añadido. Hay que citar a la empresa Quartz de Estados Unidos que utiliza una herramienta de IA llamada Quill para generar resúmenes de artículos de noticias. Quill puede, **a partir de artículos de noticias complejos, generar resúmenes concisos y fáciles de leer.** Esos resúmenes se denominan en inglés *Snabbversions* (o en castellano “versiones

rápidas") y gracias a los mismos ha aumentado la participación general. Al parecer sus lectores jóvenes son más propensos a hacer clic para obtener información adicional.

Por otra parte, mientras que en medios locales se está usando la IA para integrar directamente noticias relacionadas con el clima o la información del tráfico en tiempo real, el periódico The New York Times utiliza **una herramienta de IA para recomendar artículos a sus lectores**. Esta aplicación tiene en cuenta una gran variedad de factores, como los artículos que el lector ha leído con anterioridad y los temas que le interesan. Se trata de algo similar a lo que sucede con las recomendaciones de Netflix o de las búsquedas en Amazon.

Paralelamente **la IA se está empleando para elaborar informaciones basadas en el análisis de datos**. Diferentes herramientas permiten a los periodistas estudiar grandes volúmenes de datos para identificar patrones, tendencias y posibles historias. Esto resulta útil para investigar temas complejos o para encontrar historias que podrían pasar desapercibidas. Sirva de ejemplo el trabajo realizado por el diario británico The Guardian que utilizó IA para analizar millones de documentos judiciales. En esos documentos descubrió que miles de personas habían sido condenadas erróneamente. Esta investigación condujo a una serie de artículos que expusieron el problema y llevaron a reformas legales. Durante años, las redacciones también han utilizado la IA para extraer inferencias de conjuntos de datos masivos y para gestionar los desafíos de tareas gigantescas de gestión de contenidos e investigación, como el trabajo realizado en la cobertura de los "Papeles de Panamá" que ganó el Premio Pulitzer.

En lo que respecta a contenidos en otros formatos **los medios de forma generalizada están empleando la IA en la generación de imágenes**. Varias publicaciones hacen un uso especialmente bueno de esta tecnología. Sirvan de ejemplo el KölnerStadt-Anzeiger (Alemania) y DennikN (Eslovaquia) que emplean herramientas como Midjourney para crear ilustraciones gráficas sobre diferentes temáticas. En ambos casos, y como ocurre de forma generalizada, identifican que las imágenes han sido generadas artificialmente.

Otra aplicación de la IA en los medios es la **creación de presentadores y locutores de noticias**. En este ámbito Radio Expres en Eslovaquia ha clonado la voz de una presentadora popular Bára Hacsí cuya réplica de inteligencia artificial

(Hacsiko) ahora cubre el turno de noche, con comentarios sobre la música y noticias tomadas del sitio web de noticias de la propia compañía (Seznam-Zprávy). Por otra parte, dos estaciones de radio en el oeste de Inglaterra usan una voz sintética para convertir el texto en un boletín de radio cada hora.

Junto a la generación de audio con voces clonadas de periodistas o creadas por una IA se está llevando a cabo la generación de canales de televisión. Hay que destacar a la empresa de Johannesburgo (Sudáfrica) News-GPT (<https://newsgpt.ai/>) que ofrece un servicio de televisión experimental de 24 horas en el que todas las historias y todos los presentadores son generados por IA, sin intervención humana. Mediante un aviso, se advierte que el contenido puede contener imprecisiones o resultados inesperados. NewsGPT, que se puede ver a través de YouTube, se promociona como una empresa que ofrece noticias “sin sesgos humanos”.



Un paso más en esta dirección lo está preparando Channel.1 AI (<https://www.channel1.ai/>), con sede en Los Ángeles (Estados Unidos), quien antes de finales del 2024 pretende poner en marcha un servicio audiovisual de noticias personalizadas que aprenda qué historias quiere ver un espectador y se las ofrezca en cualquier idioma.



En otro orden de cosas **la IA se utiliza para verificar datos en artículos de noticias y redes sociales.** Esto ayuda a combatir la desinformación y garantiza que los lectores reciban información precisa. Sirva el caso de la agencia de noticias Associated Press que utiliza una herramienta llamada Fact Check para verificar automáticamente los hechos en sus artículos. Esta aplicación utiliza una variedad de técnicas, como el análisis del lenguaje natural y el reconocimiento de entidades nombradas para identificar posibles afirmaciones falsas.

En España hay varios servicios abiertos de verificación de noticias como el Verifica de la Agencia EFE (<https://verifica.efe.com/>) o Newtral (<https://www.newtral.es/zona-verificacion/fact-check/>). Por su parte, la principal organización de verificación de datos del Reino Unido, Full Fact, ha integrado una serie de técnicas de IA para ayudar a los verificadores humanos a identificar nuevas afirmaciones, incluso de transmisiones en



tiempo real, mediante la transcripción de voz a texto, comparándolas con verificaciones de datos anteriores y ayudando a verificar o desacreditar las afirmaciones.

La IA, además, se emplea en el campo de la comunicación para detectar sesgos y analizar de imparcialidad en la cobertura de noticias. Destaca el trabajo desarrollado por el Project Implicit (<https://www.projectimplicit.net>) nacido en la Universidad de Harvard en Massachusetts (EE.UU.), que, entre otras cosas, usa herramientas basadas en la IA para analizar artículos de noticias en busca de sesgos de lenguaje. La herramienta puede identificar palabras y frases que están asociadas con ciertos grupos de personas o ideas.



Otra innovación relacionada con la IA es la creación de Chatbots de atención al cliente. Se trata de automatismos destinados a proporcionar soporte al lector, oyente o espectador en los sitios web y en las aplicaciones de noticias. Estos pueden ayudar a responder preguntas de los lectores y a resolver problemas técnicos. A modo de ejemplo sirva el del periódico The New York Times que utiliza un chatbot de IA para responder preguntas de los lectores sobre sus suscripciones y otros temas relacionados con el medio de comunicación.

También la IA se utiliza para crear visualizaciones de datos interactivas y para crear experiencias de realidad aumentada y virtual que permiten a los lectores sumergirse en las noticias. De esta forma se ayuda a los lectores a comprender información compleja de manera más fácil. Por ejemplo, el periódico Los Ángeles Times utilizó IA para crear una experiencia de realidad virtual que permitió a su audiencia a ver cómo sería estar dentro de un incendio forestal.

Todos estos avances relacionados con la IA y los medios de comunicación han generado diversas críticas. Hay quien considera que se está exponiendo contenido generado por IA a los usuarios sin reglas claras sobre el etiquetado o sin asumir los riesgos de que la tecnología cometa errores (las llamadas alucinaciones). Durante 2023, se descubrió que los artículos generados automáticamente por CNET estaban plagados de errores. También que SportsIllustrated (SI) había incluido reseñas de productos de un tercero que supuestamente estaban escritas, al menos en parte, por IA, sin informar adecuadamente.

Según el informe anual titulado "Tendencias y predicciones del periodismo, los medios y la tecnología para 2024" (<https://reutersinstitute.politics.ox.ac.uk/journalism-media-and-technology-trends-and-predictions-2024>) del Instituto Reuters para el Estudio del Periodismo, "alrededor del 70% de los editores senior y directores ejecutivos de empresas de medios creen que la IA y la IA generativa limitarán la confianza general del público en las noticias".



Por su parte, Craig Forman, presidente del Centro de Noticias, Tecnología e Innovación (CNTI), en una entrevista realizada en mayo del 2024 por Eléonore de Vulpillières para la revista de Sciences Po afirmaba que: "La falta de conciencia de la audiencia sobre la variedad de usos de la IA en las noticias, incluidas aquellas que son beneficiosas para hacer llegar al público noticias de alta calidad de manera oportuna, ha influido negativamente en la confianza del público en el contenido de las noticias" (<https://www.sciencespo.fr/public/chaire-numerique/en/2024/05/06/interview-what-impact-will-ai-have-on-media-interview-with-craig-forman/>).



Así que, a pesar de los numerosos beneficios, la integración generalizada de la IA en la industria de los medios plantea desafíos y consideraciones éticas. Algunas están relacionadas con la calidad del contenido, la privacidad, el sesgo en los algoritmos, los derechos de autor, la desinformación, la necesidad de mantener estándares éticos en la redacción automatizada y el uso responsable de la IA en la comunicación. Otras con la personalización extrema y las burbujas de filtros. Todas son más o menos graves y están afectando a la forma en que nuestras sociedades funcionan y comparten ideas.

Por todo ello, y tal como se está debatiendo en la Comisión de IA de la Federación de Asociaciones de Periodistas de España FAPE -de la que forma parte la autora de estas líneas-, es necesario que los profesionales de los medios de comunicación, las empresas tecnológicas, los investigadores y los reguladores trabajen juntos para garantizar un buen uso de la IA. Y que lo hagan a través de políticas legales u otros tipos de normas para mitigar los

riesgos de los *deepfakes* y favorecer la integridad, la diversidad y la fiabilidad de la información.

En esta línea el Acuerdo Tecnológico para Combatir el Uso Engañoso de la Inteligencia Artificial en las Elecciones de 2024, firmado por las principales empresas tecnológicas, describe los compromisos para mitigar los riesgos de la IA. Estos incluyen el desarrollo de tecnología para detectar y abordar el contenido engañoso, el fomento de la resiliencia en todo el sector y la mejora de la transparencia pública y la alfabetización mediática. 

IMPACTO EN LA SOCIEDAD

💬 José Ortega

SECTOR DE LA SALUD

Diagnóstico y Tratamiento

- **Diagnóstico por Imagen:** Herramientas de IA pueden analizar imágenes médicas, como radiografías y resonancias magnéticas, con alta precisión, ayudando a detectar enfermedades como el cáncer en etapas tempranas. **Empresa:** IBM Watson Health, Google Health, Siemens Healthineers, GE Healthcare.
- **Análisis de Datos Clínicos:** Los algoritmos de IA pueden analizar grandes volúmenes de datos médicos para identificar patrones que los humanos podrían pasar por alto, ayudando en la personalización de tratamientos. **Empresa:** IBM Watson Health.

Tomografía Axial Computarizada (TAC)

- **Detecta anomalías** en las imágenes que podrían pasar desapercibidas para los radiólogos. Además, procesa y analiza grandes volúmenes de datos de imágenes mucho más rápido que los humanos, lo que acelera el tiempo de diagnóstico. **Empresa:** IBM Watson Health.
- **Precisión Diagnóstica:** Los algoritmos de IA evaluando el estadio de una enfermedad, lo que ayuda a los médicos a decidir el mejor curso de tratamiento. **Empresa:** Siemens Healthineers, GE Healthcare
- **Evaluación de la Respuesta al Tratamiento:** La IA puede analizar series de imágenes TAC a lo largo del tiempo para evaluar cómo responde un tumor o una enfermedad al tratamiento. **Google Health.**
- **Planificación de Cirugías y Radioterapia:** En oncología, la IA puede ayudar a delinear con precisión los límites de los tumores, asistiendo en la planificación de intervenciones quirúrgicas o tratamientos de radioterapia con mayor exactitud.

Ejemplos de Aplicaciones Específicas del TAC

- **Detección de Nódulos Pulmonares:** Los sistemas de IA pueden detectar nódulos en los pulmones que podrían ser indicativos de cáncer, incluso en fases muy tempranas.
- **Evaluación de Enfermedad Pulmonar:** En el contexto de enfermedades como el COVID-19, la IA ha sido utilizada para evaluar la extensión de la neumonía en los pulmones.
- **Detección de Hemorragias y Fracturas:** En casos de trauma, la IA puede identificar rápidamente hemorragias cerebrales, fracturas y otras lesiones críticas en las imágenes de TAC, facilitando la toma de decisiones clínicas urgentes.

Medicina Personalizada

- **Terapias Personalizadas:** La IA permite el análisis de datos genéticos y médicos para personalizar tratamientos, lo que puede mejorar la efectividad de los medicamentos y reducir efectos secundarios. **Empresa:** Google Health, Roche, Illumina.
- **Gestión de Recursos:** Algoritmos de IA optimizan la gestión de inventarios, la asignación de personal y la programación de citas.
- **Asistentes Virtuales y Chatbots:** Se utilizan para triaje inicial, recordatorios de medicación y citas.
- **Descubrimiento de Fármacos:** La IA acelera el proceso de descubrimiento de nuevos medicamentos. **Empresa:** Siemens Healthineers, Roche, Johnson & Johnson.
- **Vacunas:** Puede analizar secuencias genéticas de patógenos para identificar proteínas clave acelerando así el desarrollo de nuevas vacunas. **Empresa:** Pfizer.
- **Optimización de Fórmulas:** Algoritmos de aprendizaje automático ayudan a optimizar las fórmulas de las vacunas para maximizar la respuesta inmunitaria y minimizar efectos adversos. **Empresa:** Pfizer.
- **Modelos Predictivos:** La IA puede predecir la eficacia y posibles efectos secundarios de una vacuna en diferentes poblaciones, ayudando a diseñar ensayos clínicos más efectivos y seguros.

Tratamientos Contra el Cáncer

- **Screening de Fármacos:** La IA permite analizar grandes bibliotecas de compuestos para identificar nuevas moléculas con potencial terapéutico, acelerando el descubrimiento de nuevos tratamientos contra el cáncer.
- **Rediseño de Medicamentos:** Algoritmos de IA pueden sugerir modificaciones en medicamentos existentes para mejorar su eficacia o reducir toxicidades.
- **Terapias Dirigidas:** Utilizando datos genómicos y moleculares, la IA ayuda a desarrollar terapias dirigidas que atacan específicamente las mutaciones o características moleculares presentes en los tumores de un paciente.
Empresa: IBM Watson Health.

Administración de Antibióticos

- **Diagnóstico Rápido:** La IA puede ayudar a identificar infecciones bacterianas más rápidamente, permitiendo un uso más adecuado y dirigido de los antibióticos, lo que es crucial para combatir la resistencia antimicrobiana.
- **Recomendaciones de Tratamiento:** Sistemas de IA pueden sugerir los antibióticos más efectivos basándose en patrones de resistencia local y características del paciente, mejorando la administración y reduciendo el uso innecesario de antibióticos.
- **Vigilancia de Datos:** Algoritmos de IA analizan datos de salud pública y hospitalaria para detectar y predecir patrones de resistencia antimicrobiana, ayudando en la formulación de políticas de salud pública.

Inconvenientes y Desafíos

- **Privacidad y Seguridad de Datos:** El uso de grandes volúmenes de datos personales y médicos plantea preocupaciones sobre la privacidad y la seguridad, por la posibilidad de fugas de datos.
- **Desigualdad en tratamientos:** La integración de tecnologías avanzadas puede aumentar la brecha en el acceso a cuidados de salud, ya que no todas las instituciones o regiones tienen la infraestructura necesaria para implementarlas.

- **Dependencia Tecnológica:** La creciente dependencia de sistemas de IA puede llevar a la deshumanización del cuidado y a la reducción de la autonomía de los profesionales de la salud.
- **Bias y Discriminación:** Los algoritmos de IA pueden perpetuar sesgos si están entrenados en datos que no representan adecuadamente la diversidad de la población, lo que podría llevar a diagnósticos o tratamientos inadecuados.
- **Cuestiones Éticas y Regulatorias:** La implementación de IA en la salud plantea cuestiones éticas sobre el consentimiento informado, la transparencia de los algoritmos y la responsabilidad en caso de errores en el diagnóstico o tratamiento.
- **Situación Actual y Futuro:** La IA está cada vez más integrada en el sector sanitario, con muchos proyectos piloto y aplicaciones en uso clínico. Sin embargo, la adopción generalizada enfrenta desafíos relacionados con la regulación, la validación de tecnologías y la integración con sistemas existentes. Se espera que en el futuro, la IA siga transformando el sector, especialmente con el avance de tecnologías como la medicina personalizada, la telemedicina y la gestión de la salud a nivel poblacional. **Empresa:** Philips Healthcare.

SECTOR DE LA EDUCACIÓN

Personalización del Aprendizaje

- **Aprendizaje Adaptativo:** Se puede crear experiencias de aprendizaje personalizadas al ajustar el contenido y el ritmo de enseñanza según las necesidades y el progreso de cada estudiante. **Empresa:** Coursera, Duolingo.
- **Asistentes Virtuales:** Los chatbots y asistentes virtuales pueden proporcionar soporte 24/7 a los estudiantes, respondiendo preguntas, ofreciendo tutoría y ayudando en la resolución de problemas.

Acceso Ampliado a la Educación

- **Plataformas de Aprendizaje Online:** La IA permite el desarrollo de plataformas de aprendizaje online más interactivas y efectivas, accediendo a la educación a personas en diferentes ubicaciones geográficas o con limitaciones de tiempo.

- **Traducción y Accesibilidad:** Se puede traducir automáticamente materiales educativos y crear versiones accesibles para personas con discapacidades, promoviendo una educación más inclusiva.

Eficiencia Administrativa

- **Automatización de Tareas Administrativas:** Automatizar tareas como la inscripción de estudiantes, la gestión de expedientes académicos y la planificación de horarios.
- **Análisis de Datos Educativos:** Las instituciones pueden utilizar IA para analizar datos de rendimiento académico, identificar tendencias y mejorar la calidad de la educación.

Evaluación y Feedback Inmediato

- **Corrección Automatizada:** Como corregir exámenes y tareas automáticamente, proporcionando a los estudiantes un feedback inmediato y detallado sobre su desempeño.
- **Identificación de Necesidades de Apoyo:** Se puede analizar datos de rendimiento para identificar estudiantes en riesgo de rezagarse y sugerir intervenciones tempranas, como tutorías o recursos adicionales. **Empresa:** Duolingo
- **Predicción de Tendencias:** Para predecir tendencias en la matriculación, la elección de cursos y la demanda laboral futura, ayudando a adaptar los currículos a las necesidades cambiantes del mercado laboral.
- **Detección de Plagio y Calidad de Contenido:** Herramientas de IA ayudan a detectar plagio y a evaluar la calidad del contenido educativo, asegurando la integridad académica.
- **Mejora de la Experiencia de Aprendizaje:** Simulaciones y Realidad Virtual, permitiendo a los estudiantes experimentar situaciones prácticas y realistas, como experimentos de laboratorio o exploraciones históricas. **Empresa:** Knewton
- **Aprendizaje de Idiomas:** Aplicaciones de IA pueden ayudar en el aprendizaje de idiomas mediante la corrección de pronunciación, gramática y la práctica de conversación en tiempo real. **Empresa:** Carnegie Learning.

- **Desarrollo de Competencias Digitales:** El uso de herramientas de IA en el aula ayuda a los estudiantes a desarrollar competencias digitales y tecnológicas.

Inconvenientes y Desafíos

- **Desigualdad en el Acceso a la Tecnología.** La dependencia de tecnologías avanzadas puede aumentar la brecha educativa entre estudiantes con diferentes niveles de acceso.
- **Privacidad y Seguridad de Datos:** La recopilación y análisis de datos personales y académicos de los estudiantes plantea preocupaciones sobre la privacidad y la seguridad de la información.
- **Dependencia Tecnológica y Pérdida de Habilidades Humanas:** La sobredependencia en herramientas de IA puede reducir la capacidad de los estudiantes para desarrollar habilidades como el pensamiento crítico y la resolución de problemas sin asistencia tecnológica. Además existe el riesgo de deshumanizar el proceso educativo.
- **Calidad y Sesgo de los Algoritmos:** Los algoritmos de IA pueden tener sesgos inherentes si están entrenados en datos que no representan adecuadamente la diversidad de la población estudiantil.
- **Resistencia al Cambio:** La adopción de tecnologías de IA puede encontrar resistencia entre educadores y administradores que prefieren métodos tradicionales de enseñanza y gestión. 📄

INTELIGENCIA ARTIFICIAL GENERATIVA EN LA EDUCACIÓN SUPERIOR

por Ignacio Despujol

EL CAMBIO INMINENTE

Vivimos en tiempos de cambio acelerado en los que la introducción de las herramientas de inteligencia artificial generativa hará temblar los cimientos de la educación superior a muy corto plazo.

El lanzamiento de ChatGPT a finales de 2022 puso directamente en manos de los usuarios, a través de una sencilla interfaz conversacional, capacidades que solo un par de años antes parecían imposibles de alcanzar. De repente, desde estudiantes hasta profesionales pudieron aprovechar las capacidades de vanguardia de la IA generativa integradas de forma fluida en su trabajo.

En menos de un año, herramientas como ChatGPT, Claude, DALL-E o Stable Diffusion, entre muchas otras, han hecho que, aunque un buen número de profesores todavía no se ha dado cuenta, muchas de las formas de evaluar utilizadas en educación superior hayan quedado obsoletas. Estos asistentes de IA pueden generar texto, imágenes, código y más de forma similar a los humanos sobre prácticamente cualquier tema, transformando por completo el panorama educativo.

En el entorno educativo actual, las herramientas de IA generativa se están volviendo omnipresentes entre los estudiantes, que suelen ir por delante de los profesores en lo que a tecnología se refiere, y que ya las están utilizando para resolver las tareas y evaluaciones que les asignan sus profesores, llegando a niveles que han hecho que ya se haya publicado en alguna revista educativa en los Estados Unidos un artículo con un título tan gráfico como es “La muerte del trabajo para casa” (<https://dataworks-ed.com/blog/2023/07/the-death-of-homework/>).



Pero intentar prohibir su uso en el aula tiene tan poco sentido como la rebelión de los luditas contra las máquinas de vapor en el siglo XIX, ya que,

aunque podamos llegar a conseguirlo (algo que cada vez se parece más a ponerle puertas al campo), estaríamos privando a nuestros estudiantes del acceso a unas herramientas de productividad impresionante que sí conocerán los egresados de aquellos centros de enseñanza que las hayan integrado de forma adecuada.

El potencial de estas herramientas va mucho más allá de asistir con las tareas académicas y exámenes. La IA generativa ha generado un cambio de paradigma que exige que reevaluemos qué y cómo enseñamos para preparar verdaderamente a los estudiantes para un mundo impulsado por la IA.

Muchos profesores y centros han comenzado a explorar cómo aprovecharlas de manera eficaz para mejorar la productividad, potenciar la investigación y enriquecer de la experiencia de aprendizaje.

CÓMO PUEDEN USAR LA INTELIGENCIA ARTIFICIAL GENERATIVA LOS ESTUDIANTES

La IAG puede ser utilizada por los estudiantes de múltiples formas para incrementar su productividad. Comentaremos en este apartado las más generales, pues la potencia y versatilidad de esta herramienta hace imposible enumerarlas todas:

- Asistentes para tareas de escritura y codificación: Los estudiantes pueden obtener retroalimentación, sugerencias, explicaciones y hasta borradores iniciales de la IA para asistir con ensayos, informes, soluciones matemáticas y proyectos de codificación.
- Tutores socráticos: la IA puede proporcionar conocimientos complementarios para ampliar los conceptos de las lecciones, ofrecer explicaciones alternativas, generar ejemplos/analogías y responder preguntas de seguimiento. Como ejemplo tenemos la web educativa Khan Academy que ofrece un tutor socrático llamado Khanmigo (<https://www.khanmigo.ai/>).
- Asistentes para la automatización de tareas repetitivas: Las capacidades generativas simples pueden automatizar formateo, procesamiento de datos y otras tareas repetitivas que antes requerían esfuerzo del instructor/estudiante.

- **Recopiladores de información:** Las herramientas de IA pueden recopilar información de muchas fuentes de forma muy rápida y resumirla, estructurarla y formatearla eficazmente.
- **Exploración creativa:** la IA generativa permite la ideación y exploración rápidas de conceptos creativos a través de texto, imágenes, audio, animaciones y otros medios.

Es importante reseñar que, aun siendo muy poderosos, estos asistentes de IA todavía tienen limitaciones significativas. Los algoritmos que los soportan tienen tendencia a inventarse información si no disponen de ella y amplifican sesgos e incorrecciones de forma no intencionada. También tienen problemas con la originalidad, la ética y el razonamiento complejo más allá de sus datos de entrenamiento. Por tanto, la supervisión humana y una sólida instrucción sobre el uso efectivo de la IA son esenciales para sacarles el mayor partido.

La IA no va a sustituir al ser humano (al menos en el futuro cercano), va a aumentar exponencialmente sus capacidades y su productividad, con lo que se creará una gran brecha entre quien sepa utilizarla adecuadamente y quien no sepa.

USOS DE LA IAG PARA REDEFINIR LA EDUCACIÓN SUPERIOR

Estas herramientas no solo potencian las capacidades de los estudiantes, también las de los profesores y administradores de los centros de educación superior, existiendo multitud de aplicaciones que pueden contribuir a mejorar y democratizar en gran medida la educación superior. La IAG puede utilizarse para:

1. La personalización del Aprendizaje

La IAG permite crear experiencias de aprendizaje altamente personalizadas. Puede adaptar el contenido, el ritmo y el estilo de enseñanza a las necesidades individuales de cada estudiante. Por ejemplo:

- Generación de materiales de estudio adaptados al nivel y estilo de aprendizaje de cada alumno.
- Creación de ejercicios y problemas personalizados que se ajustan dinámicamente a medida que el estudiante progresa.

- Sistemas de tutoría virtual que pueden proporcionar retroalimentación instantánea y orientación personalizada.
- 2. La asistencia en la Investigación
En el ámbito de la investigación académica, la IAG está demostrando ser una herramienta poderosa que permite:
 - Análisis y síntesis de grandes volúmenes de literatura científica.
 - Generación de hipótesis y sugerencias para nuevas direcciones de investigación.
 - Asistencia en la redacción y edición de artículos académicos.
- 3. La creación de Contenido Educativo
La IAG está revolucionando la creación de materiales educativos, ya que simplifica mucho la realización de tareas como:
 - Generación de libros de texto interactivos que se actualizan automáticamente con la información más reciente.
 - Creación de simulaciones y entornos virtuales para el aprendizaje práctico.
 - Producción de videos educativos y presentaciones multimedia personalizadas.
- 4. La evaluación y Retroalimentación
En el campo de la evaluación del aprendizaje la IAG facilita la creación de:
 - Sistemas de calificación automática que pueden evaluar ensayos y respuestas abiertas.
 - Generación de retroalimentación automática detallada y constructiva para los trabajos de los estudiantes.
 - Identificación temprana de estudiantes que pueden necesitar apoyo adicional.
- 5. La Administración y Gestión Institucional
En el ámbito administrativo, la IAG ofrece soluciones innovadoras:
 - Optimización de horarios y asignación de recursos.
 - Predicción de tendencias de inscripción y planificación estratégica.
 - Automatización de tareas administrativas rutinarias.
 - Creación de sistemas automáticos de información y resolución de dudas.

LA TRANSFORMACIÓN DE LA EDUCACIÓN QUE VIENE

Los casos de uso actuales de la IA generativa están empezando a cambiar el paradigma de la educación superior, y no son más que la punta del iceberg

de lo que se podrá hacer (la aparición de los primeros sistemas multimodales, capaces de trabajar con texto, vídeo y audio simultáneamente, por ejemplo, abre un nuevo campo de posibilidades). Con los cambios disruptivos que la IAG traerá, se hace necesaria evolución de las filosofías educativas, los planes de estudio y los objetivos docentes para alinearse con una realidad abundante en IA. La capacidad de la IA para automatizar muchas tareas e impulsar una rápida exploración creativa hará que varias áreas tradicionales de instrucción queden obsoletas.

Los estudiantes pasarán menos tiempo aprendiendo tareas que la IA pueda automatizar fácilmente y más tiempo desarrollando habilidades de alto nivel que la IA no pueda replicar fácilmente. Este cambio llevará a los educadores a reevaluar sus prioridades y enfoques.

Una implicación importante es que la memorización de procedimientos, hechos y ejecución de tareas disminuirá su importancia en la mayoría de las disciplinas. Algo que ya estaba presente con el advenimiento de los ordenadores y la disponibilidad de información en Internet, pero que la IAG lleva a otro nivel.

Si bien construir conocimientos fundamentales básicos sigue siendo importante, la IA generativa reduce significativamente la necesidad de que los estudiantes dediquen mucho tiempo a la memorización rutinaria de detalles que los asistentes de IA pueden recuperar sin esfuerzo. El poner más énfasis en comprender los principios básicos, razonar sobre la información y aplicar el conocimiento de manera creativa es algo en lo que los sistemas educativos modernos llevan insistiendo ya varias décadas, pero que ahora se hace imperativo abordar de forma inmediata.

Con la IA manejando los procedimientos rutinarios y la recuperación de conocimientos, los estudiantes podrán dedicar más tiempo a desarrollar habilidades únicas de los humanos que la IA todavía no domina, como la resolución creativa y empírica de problemas, el análisis crítico, el razonamiento no estructurado, la innovación con información limitada y la exploración de ideas radicalmente nuevas.

Los educadores deberán orientar sus esfuerzos hacia el desarrollo de la capacidad de los estudiantes para enmarcar adecuadamente los desafíos, evaluar racionalmente las soluciones desde múltiples ángulos e iterar

continuamente a través del proceso de resolución de problemas mientras aprovechan los asistentes de IA.

Para trabajar de manera efectiva junto a la IA generativa, estudiantes e instructores deberán volverse expertos “ingenieros de indicaciones” (“prompt engineers” en inglés), capaces de transmitir con precisión sus necesidades a través de instrucciones cuidadosamente elaboradas. Esta habilidad de formular indicaciones se convierte en una alfabetización básica cuando se opera con sistemas de IA potentes.

Además, evaluar la precisión, seguridad y ética de las salidas de un asistente de IA requiere habilidades únicas en supervisión humana, validación y metarrazonamiento que no se habían priorizado antes. Los estudiantes deben aprender a pensar críticamente sobre las posibles deficiencias, puntos ciegos e incapacidad de la IA para comprender el verdadero razonamiento.

Si bien la IA generativa puede automatizar algunas tareas creativas, su uso hará que las habilidades únicas de los humanos como la inteligencia emocional, la comunicación y la colaboración con otros sean aún más esenciales.

Los planes de estudio deberán adaptarse para nutrir la creatividad de los estudiantes más allá de lo que la IA pueda proporcionar, desarrollando ideas, perspectivas y habilidades narrativas distintivas. Los proyectos grupales, presentaciones y actividades cultivarán el trabajo en equipo, el liderazgo, la dinámica interpersonal y la comunicación de ideas complejas que la IA no puede.

EL ROL DE LOS EDUCADORES

La introducción de la IAG acelera la necesidad de evolucionar el rol de los educadores, que, tal y como vienen recomendando las nuevas pedagogías, deben pasar de ser la fuente de información única de los alumnos, a ser recomendadores, entrenadores y guías inspiradores en un viaje educativo de descubrimiento y creatividad más profundos. Los educadores:

- Cultivarán mentalidades de aprendizaje, curiosidad y voluntad de explorar a través de prueba y error, fomentando la capacidad de aprendizaje autorregulado de los estudiantes.

- Seleccionarán cuidadosamente los conceptos de las lecciones que construyan una verdadera comprensión versus conocimiento memorizado de tareas.
- Desarrollarán tareas que promuevan el análisis, la iteración y la resolución colaborativa de problemas.
- Enseñarán a los estudiantes a aprovechar y gestionar de manera responsable los asistentes de IA, ayudándoles a desarrollar su capacidad de análisis crítico sobre ellos.
- Enfatizarán las habilidades interpersonales a través de actividades grupales, presentaciones y discusiones.
- Inspirarán a los estudiantes a abordar proyectos ambiciosos, generar ideas e innovaciones originales.

En última instancia, la IA generativa será un acelerante clave para los cambios transformadores que ya están en marcha para que la educación se centre en desarrollar habilidades cognitivas de orden superior, creatividad y potencial único de los humanos que la IA aún no puede igualar. En lugar de ver a la IA como un inhibidor, los educadores deberían abrazarla como un catalizador para evolucionar la forma en que abordamos el aprendizaje en una era abundante en IA.

BENEFICIOS Y OPORTUNIDADES DE LA IAG EN LA EDUCACIÓN SUPERIOR

Si se utiliza de forma adecuada la IAG permitirá:

1. Mejora de la Eficiencia y Calidad

La IAG puede liberar a los educadores y estudiantes de tareas rutinarias, permitiéndoles enfocarse en aspectos más creativos y estratégicos de la enseñanza y generar mejores contenidos y experiencias educativas.

2. Accesibilidad y Equidad

Al personalizar y abaratar el aprendizaje, la IAG puede ayudar a nivelar el campo de juego para estudiantes con diferentes antecedentes, extracciones sociales y capacidades.

3. Innovación en la Pedagogía

La tecnología abre nuevas posibilidades para métodos de enseñanza innovadores, como el aprendizaje basado en juegos y la realidad virtual/aumentada.

DESAFÍOS Y CONSIDERACIONES ÉTICAS DE LA IAG EN LA EDUCACIÓN SUPERIOR

El uso de la IAG no está exento de desafíos y problemas potenciales:

1. Privacidad y Seguridad de Datos

El uso de IAG implica el manejo de grandes cantidades de datos personales, lo que plantea preocupaciones sobre la privacidad y la seguridad.

2. Plagio y Autenticidad Académica

La facilidad con la que la IAG puede generar contenido plantea nuevos desafíos en cuanto a la integridad académica y la detección de plagio.

3. Dependencia Tecnológica

Existe el riesgo de que los estudiantes desarrollen una dependencia excesiva de la tecnología, descuidando habilidades críticas de pensamiento independiente.

4. Sesgo y Equidad

Los sistemas de IAG pueden perpetuar sesgos existentes si no se diseñan y entrenan cuidadosamente.

5. Crecimiento de la brecha digital

La IAG, si no se implementa y regula adecuadamente, corre el riesgo de ampliar la brecha digital existente en lugar de reducirla, exacerbando las desigualdades en el acceso y aprovechamiento de los recursos tecnológicos entre diferentes grupos sociales.

6. Deshumanización de la Educación

Hay preocupaciones sobre la posible pérdida del elemento humano en la educación, crucial para el desarrollo social y emocional.

IMPLEMENTACIÓN Y MEJORES PRÁCTICAS

Para una correcta implementación del uso de la IAG en educación superior se hace necesario seguir una serie de mejores prácticas de las que enumeramos a continuación algunas de las más importantes:

1. Formación del Profesorado y el estudiantado

Es crucial invertir en la formación continua del profesorado para que puedan utilizar eficazmente estas tecnologías y en la formación del estudiantado en su uso correcto.

2. Desarrollo de Políticas Claras

Las instituciones deben establecer directrices claras sobre el uso ético y apropiado de la IAG en entornos académicos.

3. Colaboración entre Tecnología y Pedagogía

Es esencial fomentar la colaboración entre expertos en tecnología y educadores para desarrollar soluciones que realmente mejoren el aprendizaje.

4. Evaluación Continua

Se deben implementar sistemas de evaluación continua para medir el impacto real de la IAG en los resultados educativos.

5. Enfoque Centrado en el Estudiante

La implementación de IAG debe siempre priorizar las necesidades y el bienestar de los estudiantes.

6. Nivelación de la brecha digital

Se deben implementar programas de alfabetización digital inclusivos accesibles para todos los estudiantes, independientemente de su contexto tecnológico o socioeconómico, incluyendo ayudas de acceso al equipamiento y cursos introductorios gratuitos, talleres prácticos, y recursos en línea.

EL FUTURO DE LA EDUCACIÓN SUPERIOR CON IAG

1. Evolución de los Roles Educativos

El papel de los educadores evolucionará, centrándose más en la orientación, la mentoría y el desarrollo de habilidades críticas y creativas.

2. Aprendizaje Permanente y Flexible

La IAG facilitará modelos de educación más flexibles y adaptables, apoyando el concepto de aprendizaje a lo largo de la vida.

3. Interdisciplinariedad y Colaboración Global

La tecnología permitirá una mayor colaboración entre disciplinas y a nivel internacional, enriqueciendo la experiencia educativa.

4. Personalización a Escala

Se espera que la personalización del aprendizaje alcance niveles sin precedentes, adaptándose no solo al estilo de aprendizaje, sino también a las aspiraciones profesionales y personales de cada estudiante.

5. Nuevos Campos de Estudio

Surgirán nuevas áreas de estudio centradas en la ética de la IA, la interacción humano-máquina y la alfabetización en IA.

CONCLUSIÓN

La integración de la inteligencia artificial generativa en la educación superior representa una oportunidad transformadora. Si bien los beneficios potenciales son enormes, desde la personalización del aprendizaje hasta la innovación en la investigación, es crucial abordar los desafíos éticos y prácticos que surgen.

El éxito en la implementación de la IAG en la educación superior dependerá de un enfoque equilibrado que aproveche las fortalezas de la tecnología mientras preserva los valores fundamentales de la educación. Esto requiere una colaboración estrecha entre educadores, tecnólogos, legisladores y estudiantes para crear un ecosistema educativo que sea innovador, ético y centrado en el ser humano.

A medida que avanzamos hacia esta nueva era, es fundamental recordar que la tecnología debe ser una herramienta para potenciar, no para reemplazar, la experiencia humana en la educación. El objetivo final debe ser crear un sistema educativo que prepare a los estudiantes no solo para utilizar la tecnología, sino también para pensar críticamente sobre ella, garantizando que estén equipados para liderar y prosperar en un mundo cada vez más moldeado por la inteligencia artificial.

La revolución de la IAG en la educación superior está comenzando apenas, y su trayectoria futura dependerá de cómo naveguemos colectivamente los desafíos y oportunidades que presenta. Con un enfoque reflexivo y ético, la IAG tiene el potencial de democratizar el acceso a una educación de alta calidad, fomentar la innovación y preparar a los estudiantes para un futuro que aún estamos imaginando.

En última instancia, el éxito de la IAG en la educación superior no se medirá solo por la sofisticación de la tecnología, sino por su capacidad para inspirar, empoderar y elevar el potencial humano. A medida que avanzamos, debemos mantener una visión clara de lo que significa una educación verdaderamente transformadora y utilizar la IAG como un medio poderoso para alcanzar ese fin.

Este capítulo ha sido redactado con ayuda de herramientas de IAG. 📄

AUDITORÍA, CONTABILIDAD Y ASESORAMIENTO FISCAL Y FINANCIERO

🗨 José Ortega

AUDITORÍA Y CONTABILIDAD

En la próxima década, la Inteligencia Artificial (IA) tendrá un impacto significativo en los sectores de auditoría y contabilidad. Este impacto se manifestará en varias áreas clave, desde la automatización de tareas rutinarias hasta la creación de nuevas profesiones y roles. A continuación, se detallan las principales formas en que la IA afectará estos sectores:

Automatización de Tareas Rutinarias

- **Contabilidad Automatizada y Robótica (RPA):** La automatización robótica de procesos (RPA) será una herramienta esencial, permitiendo que tareas repetitivas y basadas en reglas sean manejadas por bots. Esto incluirá la entrada de datos, conciliaciones bancarias, y generación de informes financieros. Los sistemas RPA evolucionarán para manejar tareas cada vez más complejas, como el cálculo de impuestos en múltiples jurisdicciones o la gestión de inventarios.
- **Reconocimiento Óptico de Caracteres (OCR) Integrado con IA:** La tecnología de OCR, potenciada por IA, se utilizará para escanear y procesar documentos físicos (facturas, contratos, recibos) de manera automática y eficiente, integrándose directamente en los sistemas contables. Esto minimizará el trabajo manual y mejorará la precisión de los registros contables.
- **Automatización del Ciclo de Vida Completo de las Transacciones:** Desde la generación de facturas hasta el pago de proveedores y el seguimiento de cuentas por cobrar, la IA puede gestionar y automatizar todo el ciclo de vida de las transacciones financieras, reduciendo la necesidad de intervención humana.

Análisis Predictivo y Toma de Decisiones

- **Algoritmos de Aprendizaje Automático para Previsiones Financieras:** Los modelos de IA avanzarán en su capacidad para analizar grandes volúmenes de datos históricos y actuales, mejorando las previsiones financieras. Estos algoritmos podrán identificar patrones ocultos y proporcionar predicciones más precisas sobre ingresos, gastos, y flujos de caja futuros.
- **Auditoría Predictiva:** Además de la auditoría continua, los sistemas de IA podrán predecir áreas de riesgo potencial antes de que ocurran. Esto permitirá a los auditores centrarse en la prevención y mitigación de riesgos, en lugar de solo detectar problemas después de que hayan ocurrido.
- **IA Generativa en Análisis Financiero:** Herramientas como las IA generativas se integrarán en el análisis financiero, proporcionando recomendaciones automáticas basadas en análisis de escenarios y simulaciones. Estas herramientas podrán sugerir estrategias de optimización fiscal o ajustes presupuestarios, facilitando la toma de decisiones.

Desplazamiento de Profesiones Tradicionales

- **Reducción en la Demanda de Técnicos Contables:** A medida que las plataformas de contabilidad se vuelvan más inteligentes y automatizadas, la necesidad de técnicos contables para la entrada de datos y la gestión de registros básicos disminuirá drásticamente. En lugar de ello, se requerirá que los profesionales se enfoquen en la supervisión y la estrategia.
- **El Papel Cambiante del Auditor:** Los auditores tradicionales, que dependen principalmente de revisiones manuales y muestreo de transacciones, verán cómo su rol se transforma. Las auditorías tradicionales, que solían ser retrospectivas, se convertirán en un proceso continuo y proactivo, impulsado por la automatización y la inteligencia de datos.

Creación de Nuevas Profesiones

- **Especialistas en IA y Analítica Financiera:** Surgirán nuevas profesiones dedicadas al desarrollo, implementación y mantenimiento de sistemas de IA en contabilidad y auditoría. Estos especialistas serán responsables de optimizar los algoritmos, entrenar modelos de IA y garantizar que las herramientas de análisis se alineen con los objetivos financieros y de cumplimiento de la empresa.
- **Auditoría de Algoritmos:** Se requerirán auditores especializados en la evaluación de algoritmos de IA, asegurando que los modelos predictivos y automatizados sean justos, transparentes y libres de sesgos. Estos profesionales también evaluarán la ciberseguridad de los sistemas de IA y garantizarán la integridad de los datos.
- **Consultores en Transformación Digital y Ética en IA:** A medida que más empresas integren IA en sus procesos contables, la demanda de consultores especializados en la transformación digital aumentará. Estos consultores guiarán a las empresas en la implementación de IA, asegurando que se cumplan las normas éticas y legales.

Ética y Gobernanza

- **Supervisión Ética de IA:** Con el creciente uso de la IA en decisiones financieras críticas, habrá un aumento en la demanda de profesionales que supervisen la ética y el cumplimiento de la IA. Esto incluirá la revisión de decisiones automatizadas para asegurar que no estén sesgadas y que se ajusten a las políticas de la empresa y regulaciones externas.
- **Transparencia y Trazabilidad de Decisiones Automatizadas:** Los reguladores comenzarán a exigir que las empresas mantengan un registro claro de cómo se toman las decisiones automatizadas. Esto llevará a la creación de roles especializados en la gestión de la trazabilidad y la transparencia de las decisiones de IA.

Novedades y Futuras Innovaciones

- **Contabilidad Cuántica:** Con la evolución de la computación cuántica, se podrían desarrollar nuevos métodos de procesamiento y análisis de datos financieros, permitiendo manejar y analizar cantidades masivas de datos a una velocidad sin precedentes. Aunque aún en etapas tempranas, la contabilidad cuántica podría redefinir cómo se realizan las previsiones y análisis financieros.
- **Blockchain Integrado con IA:** La combinación de blockchain e IA permitirá una mayor transparencia y seguridad en los registros contables y auditorías. Las transacciones serán registradas en una cadena de bloques inmutable, mientras que la IA analizará y validará estos registros en tiempo real, mejorando la confiabilidad de la información financiera.
- **Plataformas Financieras Descentralizadas (DeFi) y su Auditoría:** A medida que las plataformas DeFi ganen terreno, se necesitarán nuevos enfoques de auditoría y contabilidad. Los auditores y contadores deberán adaptarse a estas tecnologías emergentes, entendiendo y validando transacciones en redes descentralizadas.

Adaptación y Capacitación Continua

- **Educación en IA y Tecnología para Contadores y Auditores:** La formación continua será crucial para que los profesionales de la contabilidad y auditoría se mantengan relevantes. Las instituciones educativas y de formación profesional deberán adaptar sus programas para incluir cursos sobre IA, análisis de datos, ciberseguridad y blockchain.
- **Desarrollo de Competencias Humanas Complementarias:** A medida que la IA se haga cargo de las tareas técnicas, las habilidades blandas como el pensamiento crítico, la ética profesional, y la comunicación efectiva se volverán más importantes para los contadores y auditores. Los profesionales deberán desarrollar estas competencias para complementar el trabajo de las máquinas.

Conclusión

La próxima década verá una transformación radical en los sectores de auditoría y contabilidad, impulsada por la IA. Mientras que muchas tareas rutinarias serán automatizadas, liberando tiempo para que los profesionales se centren en roles más estratégicos, también emergerán nuevos desafíos y oportunidades. La capacidad para adaptarse a estas innovaciones, junto con una formación continua en tecnología y ética, será clave para los profesionales que deseen prosperar en este entorno cambiante.

ASESORAMIENTO FISCAL

La profesión de asesor fiscal experimentará cambios profundos en los próximos diez años, impulsados por la integración de la Inteligencia Artificial (IA). A medida que la tecnología se vuelva más sofisticada, muchas de las tareas tradicionales de los asesores fiscales serán automatizadas, transformando la naturaleza del trabajo y creando nuevas oportunidades. A continuación, se describe cómo cambiará la profesión, qué tareas serán asumidas por la IA, qué parte de los profesionales podría perder su empleo y qué nuevos roles surgirán.

Automatización de Tareas Tradicionales en la Asesoría Fiscal

- **Preparación de Declaraciones de Impuestos:** Uno de los cambios más significativos será la automatización de la preparación de declaraciones fiscales. Los sistemas de IA serán capaces de recopilar, analizar y procesar grandes volúmenes de datos fiscales, simplificando la preparación de declaraciones y asegurando el cumplimiento con las normativas fiscales en diferentes jurisdicciones. La IA también podrá actualizarse automáticamente con los cambios en las leyes fiscales, asegurando la precisión de las declaraciones.
- **Cálculo de Obligaciones Fiscales:** Las herramientas de IA podrán calcular automáticamente las obligaciones fiscales para empresas y particulares, teniendo en cuenta deducciones, créditos fiscales y otros factores

relevantes. Esto reducirá drásticamente la necesidad de la intervención humana en cálculos complejos.

- **Gestión de Auditorías Fiscales:** La IA puede anticipar y preparar automáticamente respuestas a auditorías fiscales, analizando patrones en los datos y generando la documentación necesaria. Esto disminuirá la carga de trabajo de los asesores fiscales en caso de auditorías.

Análisis Predictivo y Planificación Fiscal Avanzada

- **Modelos Predictivos para Planificación Fiscal:** La IA permitirá a los asesores fiscales realizar análisis predictivos para la planificación fiscal a largo plazo. Al analizar datos históricos y actuales, la IA podrá sugerir estrategias fiscales optimizadas, anticipando cambios en la legislación y condiciones económicas. Esto transformará a los asesores fiscales en estrategias que ayudan a sus clientes a tomar decisiones informadas con antelación.
- **Simulaciones Fiscales:** Las herramientas de IA podrán ejecutar simulaciones fiscales en tiempo real, permitiendo a los asesores y sus clientes evaluar el impacto de diferentes escenarios fiscales antes de tomar decisiones. Esto será particularmente útil para empresas que buscan optimizar su estructura fiscal en un entorno global cambiante.

Reducción de la Demanda de Profesionales Tradicionales

- **Disminución en la Demanda de Asesores Fiscales para Tareas Rutinarias:** A medida que las tareas básicas como la preparación de declaraciones y la entrada de datos se automaticen, habrá una menor demanda de asesores fiscales que se dediquen a estos trabajos rutinarios. Los profesionales que no se adapten a roles más estratégicos o que no adquieran nuevas competencias podrían ver sus empleos en riesgo.
- **Reducción de Asesores Especializados en Cumplimiento:** Las plataformas de IA estarán cada vez más capacitadas para garantizar el cumplimiento de normativas fiscales, reduciendo la necesidad de asesores que se especializan exclusivamente en la verificación del cumplimiento normativo.

Nuevas Profesiones y Roles Emergentes

- **Especialistas en Estrategia Fiscal Basada en IA:** Surgirán nuevos roles para profesionales que se especialicen en la interpretación de los análisis generados por IA. Estos asesores utilizarán las capacidades predictivas de la IA para diseñar estrategias fiscales personalizadas para sus clientes, optimizando su carga fiscal en función de múltiples variables.
- **Consultores de Transformación Digital para Asesoría Fiscal:** A medida que las firmas de asesoría fiscal implementen IA y otras tecnologías avanzadas, se necesitarán consultores que puedan guiar esta transición. Estos profesionales ayudarán a las empresas a integrar IA en sus procesos fiscales y garantizarán que los empleados estén capacitados para utilizar estas nuevas herramientas.
- **Arquitectos de Soluciones Integradas:** Para diseñar e implementar soluciones tecnológicas integradas que abarcan todas las funciones fiscales de una empresa. Integrar sistemas de contabilidad, software de cumplimiento normativo y herramientas de análisis de datos.
- **Audidores Fiscales Especializados en IA:** La auditoría fiscal también se verá transformada, y se requerirán auditores con conocimientos en sistemas de IA para revisar y verificar la precisión y la legalidad de las decisiones fiscales automatizadas. Estos auditores también deberán garantizar que las IA operen sin sesgos y cumplan con las normativas fiscales.
- **Detectives de Fraude y Anomalías:** Usar herramientas de IA para identificar y prevenir el fraude fiscal y otras anomalías. Implementar y supervisar sistemas de detección de fraude, analizar patrones de comportamiento fiscal inusual.
- **Asesores de Seguridad de Datos:** Asegurar que los datos fiscales de los clientes estén protegidos contra ciberataques y brechas de seguridad. Implementar medidas de ciberseguridad, supervisar la seguridad de los sistemas de datos fiscales.
- **Innovadores de Servicios Personalizados:** Ofrecer servicios fiscales altamente personalizados basados en el análisis detallado de los datos del cliente.

Utilizar IA para proporcionar asesoramiento fiscal específico para cada cliente, diseñar estrategias fiscales personalizadas.

Ética y Gobernanza en Asesoría Fiscal Automatizada

- **Supervisión Ética de la IA Fiscal:** A medida que la IA asuma más responsabilidades en la toma de decisiones fiscales, será crucial asegurar que estas decisiones sean éticas y transparentes. Profesionales especializados en la ética de la IA trabajarán para prevenir el uso indebido de la tecnología en la evasión fiscal o en la optimización agresiva que podría dañar la reputación de las empresas.
- **Regulación y Cumplimiento en Sistemas Automatizados:** Se crearán roles centrados en la regulación y el cumplimiento de la IA en la asesoría fiscal, para garantizar que las herramientas automatizadas cumplan con las normativas y no se desvíen de las prácticas legales y éticas aceptadas.

Innovaciones y Futuras Tendencias

- **Integración de Blockchain en la Asesoría Fiscal:** La tecnología blockchain, combinada con IA, permitirá una mayor transparencia y seguridad en las transacciones fiscales. Los asesores fiscales deberán familiarizarse con estas tecnologías para poder ofrecer servicios de asesoría en criptomonedas y activos digitales, y para asegurar que las transacciones registradas en blockchain sean tratadas adecuadamente en términos fiscales.
- **Computación Cuántica en la Optimización Fiscal:** Aunque aún en desarrollo, la computación cuántica podría revolucionar la optimización fiscal al permitir cálculos increíblemente complejos a una velocidad sin precedentes. Los asesores fiscales que se especialicen en esta área emergente tendrán una ventaja competitiva significativa.

Capacitación y Adaptación Continua

- **Formación Continua en IA y Tecnología:** Los asesores fiscales deberán recibir formación continua en nuevas tecnologías para mantenerse

competitivos. Esto incluirá no solo el uso de herramientas de IA, sino también la comprensión de cómo estas tecnologías afectan las regulaciones fiscales y las mejores prácticas en asesoría.

- **Desarrollo de Habilidades de Comunicación y Consultoría:** A medida que las tareas técnicas se automatizan, las habilidades interpersonales y de consultoría se volverán más importantes. Los asesores fiscales deberán enfocarse en ofrecer un valor añadido a través de un asesoramiento personalizado, comprendiendo a fondo las necesidades y metas de sus clientes.

Beneficios de Estas Nuevas Funciones

- **Mayor Eficiencia:** La automatización y el uso de IA permiten a los asesores fiscales centrarse en tareas más estratégicas y de mayor valor añadido.
- **Mejor Toma de Decisiones:** Los análisis avanzados y las predicciones basadas en datos proporcionan información más precisa para la planificación fiscal.
- **Conformidad y Seguridad Mejoradas:** La supervisión continua y las medidas avanzadas de seguridad protegen mejor contra el fraude y garantizan el cumplimiento normativo.
- **Servicios Más Personalizados:** El análisis detallado de los datos del cliente permite ofrecer soluciones fiscales a medida.
- **Adaptación a Cambios Normativos:** La capacidad de ajustar rápidamente los sistemas de IA a nuevas normativas garantiza una conformidad constante.

Desafíos

- **Capacitación y Desarrollo de Habilidades:** Los asesores fiscales necesitan desarrollar nuevas habilidades técnicas y analíticas.
- **Integración de Tecnologías:** Asegurar que las nuevas herramientas y sistemas se integren sin problemas con los procesos existentes.
- **Gestión del Cambio:** Ayudar a los clientes a adaptarse a nuevas tecnologías y procesos puede ser un desafío.
- **Protección de Datos:** Mantener la seguridad y la privacidad de los datos fiscales es crucial en un entorno cada vez más digitalizado.

Conclusión

La próxima década verá una transformación radical en la profesión de asesor fiscal, impulsada por la IA. Mientras que muchas tareas rutinarias serán automatizadas, liberando tiempo para que los profesionales se concentren en roles más estratégicos, también emergerán nuevos desafíos y oportunidades. Los profesionales que adopten estas nuevas tecnologías, se especialicen en áreas emergentes como la planificación fiscal basada en IA y se adapten a las cambiantes normativas fiscales estarán bien posicionados para prosperar en este entorno. La formación continua y la capacidad de asesorar más allá de lo técnico serán claves para el éxito en esta nueva era.

ASESOR FINANCIERO

En la próxima década, la Inteligencia Artificial (IA) tendrá un impacto transformador en el sector del asesoramiento financiero, cambiando la manera en que se gestionan las inversiones, se asesora a los clientes y se toma decisiones financieras. A medida que la IA se integre en las prácticas cotidianas, ciertas tareas tradicionales desaparecerán o serán reemplazadas, mientras que surgirán nuevas oportunidades y roles dentro del sector. A continuación, se describen los cambios clave que se anticipan:

TAREAS QUE DESAPARECERÁN O SERÁN REEMPLAZADAS POR IA

Análisis Financiero Básico

- **Automatización del Análisis de Datos:** Las herramientas de IA podrán analizar grandes volúmenes de datos financieros de manera rápida y precisa, eliminando la necesidad de analistas humanos para realizar análisis financieros básicos. Esto incluye la evaluación de estados financieros, la identificación de tendencias y la generación de informes financieros rutinarios.
- **Modelos de Valuación Automatizados:** Los modelos de IA podrán realizar valuaciones de activos y empresas de forma automatizada, utilizando datos en tiempo real y ajustando las predicciones según la información más reciente. Esto reducirá la dependencia de los analistas financieros para realizar evaluaciones estándar.

Gestión de Portafolios Tradicional

- **Gestión Automatizada de Inversiones (Robo-Advisors):** Los robo-advisors, o asesores automatizados, ya han comenzado a cambiar la forma en que se gestionan las inversiones, y su uso se expandirá en la próxima década. Estos sistemas automatizados pueden construir, gestionar y reequilibrar portafolios de inversión con poca o ninguna intervención humana, utilizando algoritmos para optimizar el rendimiento basado en el perfil de riesgo del cliente.
- **Reequilibrio de Portafolios:** El reequilibrio automático de portafolios basado en algoritmos de IA eliminará la necesidad de que los asesores financieros realicen esta tarea manualmente, optimizando la asignación de activos según las condiciones del mercado y las metas del cliente.

Asesoramiento financiero estándar

- **Generación de Recomendaciones Automatizadas:** La IA será capaz de ofrecer recomendaciones financieras personalizadas basadas en datos del cliente, como su historial financiero, metas a largo plazo, y perfil de riesgo. Esto reducirá la necesidad de asesores financieros para ofrecer consejos estandarizados, ya que los algoritmos pueden personalizar las recomendaciones de manera más precisa y eficiente.
- **Planificación Financiera Básica:** Las herramientas de planificación financiera automatizadas podrán generar planes financieros personalizados que se ajusten a las metas y circunstancias del cliente, como la jubilación, la compra de vivienda o la educación de los hijos.

NUEVAS TAREAS Y ROLES QUE SURGIRÁN.

Asesoría Financiera Estratégica y Personalizada

- **Estrategias de Inversión Personalizadas:** Con la automatización de las tareas básicas, los asesores financieros se centrarán más en la creación de estrategias de inversión personalizadas que consideren factores complejos

y multidimensionales, como las emociones del cliente, cambios en la vida personal, o consideraciones éticas y sostenibles (inversiones ESG).

- **Consultoría en Finanzas Conductuales:** La combinación de IA con el conocimiento de la psicología financiera permitirá a los asesores ofrecer un servicio más integral, ayudando a los clientes a superar sesgos cognitivos y emocionales que puedan afectar sus decisiones de inversión.

Asesoría en Inversiones Alternativas y Complejas

- **Consultores de Activos Digitales y Criptomonedas:** A medida que los activos digitales y las criptomonedas se vuelven más prevalentes, los asesores financieros especializados en este tipo de inversiones emergentes serán cada vez más demandados. Estos profesionales ofrecerán asesoramiento sobre cómo integrar estos activos en un portafolio diversificado y cómo gestionar los riesgos asociados.
- **Asesoría en Finanzas Sostenibles y ESG:** La creciente importancia de las inversiones sostenibles y la responsabilidad social corporativa (ESG) requerirá asesores especializados que puedan ayudar a los clientes a alinear sus inversiones con sus valores éticos y sostenibles, utilizando herramientas de IA para evaluar el impacto y las oportunidades en este campo.
- **Impacto en Profesionales:** Esto creará oportunidades para aquellos con habilidades tanto en finanzas como en tecnología. Los profesionales con experiencia en la intersección de estos campos serán esenciales para garantizar la integridad de los sistemas financieros basados en IA.

Adopción de Tecnología y Análisis de Datos

- **Competencias Tecnológicas:** Los asesores financieros necesitarán adquirir habilidades en el uso de herramientas de IA y análisis de datos. Esto incluye la capacidad de interpretar los resultados generados por algoritmos y utilizar plataformas automatizadas de asesoría financiera (robo-advisors).
- **Integración de IA en Servicios:** La implementación de IA permitirá a los asesores ofrecer servicios más personalizados y eficientes. Por ejemplo,

pueden utilizar modelos predictivos para anticipar las necesidades de los clientes y recomendar estrategias de inversión personalizadas.

Enfoque en la Planificación Estratégica

- **Estrategia a Largo Plazo:** La IA puede manejar muchas de las tareas operativas y de análisis, permitiendo a los asesores centrarse en la planificación financiera a largo plazo y en la estrategia de inversión. Esto incluye la gestión de patrimonio y la planificación de la jubilación.
- **Asesoría Holística:** Con más tiempo para interactuar con los clientes, los asesores pueden ofrecer un enfoque más holístico (amplio), abordando no solo inversiones, sino también planificación fiscal, gestión de riesgos y planificación de herencias.

Desarrollo de Habilidades Interpersonales

- **Habilidades Blandas:** Las habilidades interpersonales como la comunicación, la empatía y la capacidad de construir relaciones serán más importantes que nunca. A medida que las máquinas gestionan tareas técnicas, los asesores humanos deberán enfocarse en comprender profundamente las necesidades y metas de sus clientes.
- **Confianza y Transparencia:** La construcción de confianza será fundamental. Los clientes deben sentir que sus asesores están actuando en su mejor interés, especialmente cuando se utilizan tecnologías avanzadas que pueden parecer opacas.

Integración de IA y Herramientas Digitales en la Asesoría Financiera

- **Consultores en Transformación Digital:** Se requerirán profesionales que guíen a las firmas financieras en la adopción e integración de herramientas de IA y otras tecnologías avanzadas en sus procesos de asesoría. Estos consultores se asegurarán de que las firmas maximicen los beneficios de la automatización y mejoren la experiencia del cliente mediante la tecnología.

- **Desarrolladores de Soluciones de IA para Finanzas:** Con la creciente demanda de soluciones personalizadas, surgirán roles para desarrolladores especializados en crear y adaptar herramientas de IA específicamente para el sector financiero, mejorando la personalización y precisión de los servicios financieros.

Asesoramiento en Finanzas Sostenibles y ESG

- **Inversiones Sostenibles:** La demanda de asesoría en inversiones que tengan en cuenta criterios ambientales, sociales y de gobernanza (ESG) continuará creciendo. Los asesores que puedan ofrecer asesoría en esta área utilizando herramientas de IA para evaluar el impacto sostenible serán altamente solicitados.
- **Impacto en Profesionales:** Los asesores financieros que se especialicen en ESG tendrán una ventaja competitiva en un mercado que valora cada vez más la responsabilidad social corporativa. Aquellos que se mantengan en modelos de asesoría tradicionales podrían quedarse atrás.

Supervisión y Validación de Algoritmos

- **Auditores de IA:** A medida que los algoritmos de IA asumen más tareas en la gestión de inversiones y el asesoramiento, surgirá la necesidad de profesionales que supervisen y auditen estos sistemas para asegurar que funcionen de manera correcta, transparente y sin sesgos. Este rol incluirá la validación de modelos predictivos y la garantía de que las decisiones automatizadas cumplan con las regulaciones y expectativas éticas.
- **Especialistas en Optimización de Algoritmos:** Se necesitarán profesionales que trabajen en la optimización continua de los algoritmos utilizados en la gestión de inversiones y asesoría financiera, asegurando que se adapten a las cambiantes condiciones del mercado y a las necesidades individuales de los clientes.

Ética y Regulación de la IA en Finanzas

- **Especialistas en Ética Financiera y IA:** La creciente dependencia de la IA en la toma de decisiones financieras plantea cuestiones éticas que deberán ser gestionadas. Surgirán roles para especialistas en ética que se aseguren de que las soluciones de IA sean utilizadas de manera justa, sin perpetuar sesgos, y respetando los valores y la privacidad de los clientes.
- **Reguladores y Cumplimiento de Normativas en IA:** A medida que las regulaciones en torno al uso de IA en finanzas se desarrollen, habrá una creciente necesidad de profesionales que supervisen el cumplimiento de estas normativas y aseguren que las tecnologías utilizadas en la asesoría financiera cumplan con los estándares legales y éticos.

Consultoría en Finanzas Digitales y Activos Emergentes

- **Especialización en Activos Digitales:** Los asesores financieros que se especialicen en áreas emergentes, como criptomonedas, blockchain, y finanzas descentralizadas, estarán en alta demanda. Estos profesionales asesorarán a los clientes sobre cómo integrar estos activos en sus portafolios y cómo manejar los riesgos asociados.
- **Impacto en Profesionales:** Los asesores que desarrollen una especialización en nuevas clases de activos y tecnologías financieras emergentes serán valiosos, mientras que aquellos que no se adapten podrían encontrar sus habilidades obsoletas.

Adaptación y Capacitación Continua

Para adaptarse a estos cambios, los profesionales en el sector de asesoramiento financiero deberán:

- **Formarse en Tecnologías Emergentes:** Los asesores deberán adquirir conocimientos en IA, big data, blockchain y otras tecnologías que están moldeando el futuro de las finanzas.

- **Desarrollar Habilidades Consultivas:** Con la automatización de tareas técnicas, las habilidades interpersonales y de consultoría estratégica se volverán cada vez más importantes.
- **Adoptar una Mentalidad de Aprendizaje Continuo:** El entorno financiero será cada vez más dinámico, y los profesionales deberán mantenerse al día con los últimos avances tecnológicos y regulatorios para seguir siendo competitivos.
- **Gestión de la ciberseguridad y la privacidad:** Con el creciente uso de la IA y la digitalización de los servicios financieros, la ciberseguridad se convierte en una prioridad. Los asesores deben estar capacitados en la protección de datos y la gestión de riesgos cibernéticos.

Conclusión

La IA revolucionará el sector del asesoramiento financiero, automatizando muchas de las tareas rutinarias y analíticas, y dando lugar a una nueva era de servicios financieros más personalizados, estratégicos y tecnológicamente avanzados. Si bien algunas tareas y roles tradicionales desaparecerán, surgirán nuevas oportunidades para aquellos que puedan adaptarse y especializarse en las áreas emergentes, como la estrategia financiera personalizada, la auditoría de IA, las inversiones alternativas y sostenibles, y la ética en el uso de la tecnología. 📄

1. EN QUÉ PUNTO NOS ENCONTRAMOS Y POR QUÉ.

En 1997, Zbigniew Brzezinski publicaba el libro "El gran tablero mundial", donde señalaba abiertamente que Estados Unidos se había convertido en la nación hegemónica en el mundo. 20 años después, en 2017, Graham Allison publicaba "Destined for war", una obra en la que, ante el ascenso de China hasta amenazar la primacía estadounidense, su autor hace un análisis de las condiciones por las cuales se podría llegar o evitar una guerra entre ambas superpotencias¹.

¿Qué ha pasado desde entonces?

El liderazgo chino se ha ido fraguando durante las últimas décadas, como en el caso de las tierras raras², donde el "Plan nacional de investigación y desarrollo de altas tecnologías" (programa 863) de 1986 le ha llevado a un control prácticamente total de las cadenas de suministro mundial³.

Ya en 2014, reforzado con el "Plan Made in China" de 2015⁴, China lanza su estrategia para priorizar la autosuficiencia en la fabricación de

-
- 1 Para ello, parte de la llamada "trampa de Tucídides", en la que cayeron las ciudades-estado de Atenas y Esparta y que les condujo a la Guerra del Peloponeso en el siglo V a.C. La "trampa de Tucídides" aparece cuando el ascenso de una nación amenaza la hegemonía de otra, siendo el resultado más probable la guerra.
 - 2 Las tierras raras están presentes en cualquier dispositivo tecnológico de nuestra vida cotidiana, en casa, en el trabajo e incluso en nuestros bolsillos (<https://projects.research-and-innovation.ec.europa.eu/en/horizon-magazine/science-rocks-rare-earth-technology-cant-live-without>).
 - 3 Como explica Juan Manuel Chomón en su publicación "La era de las tierras raras", para que nos hagamos una idea del control chino de esa cadena de suministro, si la OPEP controla el 41% de la producción de petróleo, alcanzando su líder, Arabia Saudí, un 10% del petróleo mundial, solo China controla aproximadamente el 75% de la producción de tierras raras.
 - 4 Uno de sus objetivos es incrementar la producción local de componentes básicos a un 70% en 2025, planteando un horizonte a 2049 en el que se supere a grandes potencias

semiconductores que, aunque por ahora no existe ninguna empresa en el mundo que fabrique chips con mayor precisión que la Taiwan Semiconductor Manufacturing Company (TSMC)⁵, estas tierras raras también están presentes para mejorar el rendimiento y la eficiencia de los microchips o semiconductores, lo que demuestra cómo de atados tiene China los principales sectores estratégicos a escala global.

Pero el mayor detonante del despertar chino tuvo lugar cuando la IA y, concretamente, Alphago –aplicación informática propiedad de Google (en este caso conviene enfatizar que se trata de una empresa estadounidense)–, derrotó, primero al mejor jugador del mundo –coreano– de “Go”, y un año después, en mayo de 2017, al mejor jugador chino de “Go”, juego de mesa de estrategia chino por excelencia, probablemente con más 2.500 años de existencia y que cualquier erudito chino debía dominar como una de las cuatro formas de arte. Fue su “momento Sputnik”⁶, término acuñado cuando Rusia envió el primer satélite espacial, lo que provocó la reacción de EE.UU., desatándose la carrera espacial y la creación de la NASA.

En efecto, el 20 de julio de 2017, dos meses después de esa partida, el Consejo de Estado del Gobierno chino emitió el “Plan de desarrollo de inteligencia artificial de nueva generación”⁷, cuyo objetivo es “alcanzar el nivel líder mundial en teoría, tecnología y aplicaciones generales de inteligencia artificial para 2030, convertirse en el principal centro de innovación en inteligencia artificial del mundo, lograr resultados significativos en economía y sociedad inteligentes y sentar una base importante para convertirse en un país líder en innovación y potencia económica.”

En favor de esa posible hegemonía mundial en IA, China cuenta con grandes fortalezas para las eras de la implementación, pues dispone de fervientes

tecnológicas como Alemania, Japón o Estados Unidos. ICEX (https://www.observatoriorli.com/docs/CHINA/PLAN_2025_China.pdf).

5 Miller, C. (2023). *La guerra de los chips: La gran lucha por el dominio mundial*. Península.

6 Kai-Fu Lee (2020). *Superpotencias de la inteligencia artificial: China, Silicon Valley y el nuevo orden mundial*. Deusto.

7 https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm

empresarios y un entorno político favorable, y de los datos, en la que cuenta con un vasto y creciente mercado que ya ha superado a los Estados Unidos, factor este último que prevalece sobre la disponibilidad de los mejores científicos del mundo⁸.

Por la parte de Estados Unidos, su impulso a la innovación se remonta al siglo XIX con el liderazgo de los materiales que dieron lugar a la segunda revolución industrial, pasando por la citada carrera espacial durante la Guerra Fría, el avance de las empresas tecnológicas en la era Clinton o el renovado impulso experimentado en Silicon Valley tras la crisis de las puntocom, y todo ello sustentado por una cultura claramente favorable al emprendimiento y unas universidades muy proclives la innovación aplicada en plena sintonía con el sector empresarial. Así, en 2023, 9 de las 10 primeras empresas de mundo por capitalización bursátil eran estadounidenses^{9,10}.

Como resultado, si repasamos la evolución de ambas potencias en el peso económico mundial, se aprecia como China ha pasado de menos del 4% del PIB mundial en 2000, al 9% y 18% en 2010 y 2023 respectivamente. Por su parte, Estados Unidos ha pasado en esas mismas fechas del 30% al 22% y 26% en 2023¹¹, cifras que justifican la publicación de Graham Allison con la que arrancaba este capítulo.

8 Kai-Fu Lee, op. cit.

9 <https://www.pwc.co.uk/audit/assets/pdf/global-100/global-top-100-companies-2024.pdf>.

10 Para comprender la influencia de estas grandes plataformas, el valor de mercado de las primeras empresas tecnológicas del mundo supera (o incluso duplica) el PIB de España, lo que plantea un nuevo escenario geopolítico que se superpone al tablero de competencia entre países o bloques.

En este aspecto, como afirman Kissinger, Schmidt y Huttenlocher en "La era de la inteligencia artificial", aunque los gobiernos pueden tratar de restringir el uso de estas plataformas para evitar que superen a sus rivales nacionales en regiones importantes, como los gobiernos no suelen crear ni gestionar estas plataformas, este escenario estratégico es especialmente dinámico y difícil de predecir.

11 https://datos.bancomundial.org/indicador/NY.GDP.MKTP.CD?end=2023&name_desc=false&skipRedirection=true&start=1960&view=chart. PIB en dólares a precios actuales.

Es de reseñar que, en 2023, con la medición del PIB en paridad de poder adquisitivo -dólares a precios internacionales constantes de 2017-, China habría superado a Estados

En este marco, ¿cómo ha evolucionado la Unión Europea y qué políticas está aplicando en materia de IA?

Sin duda, fue la crisis de las puntocom la que abrió la gran brecha entre EE.UU. y China con respecto a la Unión Europea. De hecho, en 2006, la financiación en fondos privados en webs 2.0 en Estados Unidos ascendió a 700 M\$, mientras que en la Unión Europea apenas alcanzó los 100 M\$. Pero, no solo se abandonaron las inversiones en el sector tecnológico, sino que, en muchos casos, este escepticismo derivó en un letargo el proceso de digitalización. Por el contrario, en Estados Unidos los casos de éxitos se multiplicaron y China copiaba¹² y mejoraba sus ideas para su implantación en su territorio. De este modo, Europa pasó a ser irrelevante en este mercado¹³, y ahora ha centrado su atención reguladora en las condiciones de participación de los operadores globales de plataformas de redes en su mercado ante su práctica inexistencia interna¹⁴.

Nuevamente en términos de PIB mundial, Europa ha pasado del 21% en el año 2000 –por encima de Estados Unidos en ese año–, al 17% y el 15% en 2010 y 2023 respectivamente.

Unidos con un 18% frente a un 15%. https://datos.bancomundial.org/indicador/NY.GDP.MKTP.PP.KD?end=2023&name_desc=false&skipRedirection=true&start=1960&view=chart

- 12 Aunque desde una perspectiva occidental la copia está estigmatizada, como se encarga de recordarnos Kai-Ku Lee en su obra sobre las Superpotencias de la IA, el enfoque chino tiene profundas raíces históricas y culturales: “Mientras Sócrates animaba a sus estudiantes a buscar la verdad cuestionándolo todo, los antiguos filósofos chinos aconsejaban a la gente que siguieran los rituales del pasado antiguo. La copia rigurosa de la perfección era percibida como la ruta hacia la verdadera maestría y durante milenios, la memorización ha constituido el núcleo duro de la educación china”. Así, aun hoy, “Las Analectas” de Confucio es un texto básico en la enseñanza secundaria china; todos los adolescentes chinos leen, estudian y memorizan máximas de este libro Castellón, J. J. (2023). *Confucio: maestro de moral, filósofo de ética*. Isidorianum).
- 13 Moreno, L., y Pedreño, A. (2020). Europa frente a EE.UU. y China. Prevenir el declive en la era de la inteligencia artificial. Publicación independiente.
- 14 Como señalan Kissinger, Schmidt y Huttenlocher en su publicación “La era de la inteligencia artificial”.

2. LA RUPTURA DE LAS CADENAS DE SUMINISTROS POSCOVID Y LA GUERRA DE UCRANIA: ¿EL MOMENTO SPUTNIK PARA LA UNIÓN EUROPEA?

A la par que se están haciendo esfuerzos para recuperar el terreno perdido como en el caso de la “Ley Europea de Chips”¹⁵, se acaba de aprobar el primer “Reglamento de Inteligencia Artificial”¹⁶ de este tipo en el mundo¹⁷:

“El Consejo ha adoptado hoy un acto legislativo innovador encaminado a armonizar las normas sobre inteligencia artificial, el denominado Reglamento de Inteligencia Artificial. Este acto legislativo emblemático adopta un enfoque basado en el riesgo, lo que significa que cuanto mayor sea el riesgo de causar daños a la sociedad, más estrictas serán las normas. Es el primero de este tipo en el mundo, por lo que puede convertirse en un referente mundial por lo que respecta a la regulación de la IA.”

Cabe reflexionar sobre este reglamento desde distintos puntos de vista:

En primer lugar, por qué adopta este enfoque basado en el riesgo¹⁸ y presume de haber aprobado el primer reglamento del mundo de este tipo.

15 <https://digital-strategy.ec.europa.eu/es/policies/european-chips-act>.

En el argumentario que justifica esta ley se puede leer: “La reciente escasez mundial de chips ha interrumpido las cadenas de suministro, ha causado escasez de productos que van desde automóviles hasta dispositivos médicos y, en algunos casos, incluso ha obligado a cerrar fábricas.” Específicamente, “la Ley Europea de Chips reforzará el ecosistema de semiconductores en la UE, garantizará la resiliencia de las cadenas de suministro y reducirá las dependencias externas. Es un paso clave para la soberanía tecnológica de la UE. Además, garantizará que Europa cumpla su objetivo de la década digital de duplicar su cuota de mercado mundial en semiconductores al 20%.” Recordemos que ya el “Plan Made in China” de 2015 planteaba la autosuficiencia en sus cadenas de suministro en un 70% en 2025, incluyendo los chips por su importancia para la seguridad nacional.

16 https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=OJ:L_202401689.

17 <https://www.consilium.europa.eu/es/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/>

18 Dando continuidad al documento de “Directrices éticas para una IA fiable” elaborado por el “Grupo de expertos de alto nivel sobre inteligencia artificial” y publicado en 2019. Concretamente, señala que la IA fiable tiene tres componentes: 1) debe ser lícita, es decir, cumplir todas las leyes y reglamentos aplicables; 2) ha de ser ética, demostrando el respeto y garantizando el cumplimiento de los principios y valores éticos, y 3) debe ser robusta, tanto desde el punto de vista técnico como social, puesto que los sistemas de IA, incluso

Obviamente, esta orientación encuentra su explicación en nuestros arraigados valores humanísticos, los cuales creo que merece la pena recordar, aunque sea de forma sucinta, como hace Fernando García de Cortázar en su obra “Érase una vez Europa”¹⁹: Europa es hija de Grecia y de Roma, del sermón de la montaña y de Pablo de Tarso, quien pronunció el primer discurso igualitario de la historia y la memorable “Primera carta a los corintios”. Ya en nuestros días, esa histórica preocupación por los derechos humanos, impregnada por episodios como el señalamiento de los judíos durante la Segunda Guerra Mundial, es la que subyace tras el Reglamento General de Protección de Datos²⁰, aprobado en 2016 y en vigor desde 2018.

Pero también Europa hunde sus raíces en el derecho romano y, como de nuevo nos recuerda Fernando García de Cortázar, en una “democracia basada en los valores fríos del derecho, de ahí que los europeos debamos mucho a los juristas de Bolonia, que junto a los traductores de Toledo y los teólogos que aprendieron a dudar con Abelardo en París, ayudaron a poner los cimientos de las sociedades prósperas y libres que hoy tenemos en Occidente”.

Es decir, Europa encuentra también en su génesis una gran tradición jurídica, a diferencia de China, donde la ciencia jurídica no ha jugado históricamente un gran papel: “el hombre debe tratar de vivir en armonía con el universo y con la sociedad a la que pertenece; por tanto, el buen ciudadano es el hombre de compromisos, y el insistir en pretendidos derechos es una inmoralidad”²¹.

Por otro lado, estas tesis regulatorias de la Unión Europea casan con el llamado “Efecto Bruselas”²², cuya autora, Anu Bradford, sostiene que, a pesar de la pérdida de influencia económica de la UE durante los últimas décadas, las características intrínsecas del mercado interior de la UE y de determinados

si las intenciones son buenas, pueden provocar daños accidentales. La fiabilidad de la IA no concierne únicamente a la fiabilidad del propio sistema de inteligencia artificial, sino también a la de todos los procesos y agentes implicados en el ciclo de vida del sistema.

19 García, F. (2023). *Érase una vez Europa: Senderos de justicia, tolerancia y libertad*. Espasa.

20 <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A32016R0679>.

21 Floris, G. (2022). *Panorama de la historia universal del derecho*. Marcial Pons.

22 Bradford, A. (2020). *The Brussels Effect: How the European Union Rules the World*. Oxford University Press.

mercados digitales como los que aquí se tratan, propician que el marco regulatorio europeo sea adoptado por terceros países, por lo que la influencia de la UE seguirá siendo muy significativa globalmente, incluidos Estados Unidos y China, desplegando, de esta forma, nuestros valores occidentales.

Desde mi punto de vista, de estas afirmaciones se derivan, al menos, dos cuestiones fundamentales:

1. ¿Será realmente Europa capaz de ejercer esa influencia internacional?

Si bien la autora trata de justificar por qué seguirá haciéndolo, me parece que, si continúa la pérdida de relevancia económica de la UE como lo ha hecho durante los últimos veinte años, también lo hará esa influencia porque, precisamente, cuanto menor sea ese peso, menos relevancia tendrá el mercado europeo para las grandes plataformas.

2. Aun cuando se mantuviese esa influencia regulatoria: ¿favorece eso que Europa vuelva a recuperar posiciones de vanguardia perdidas?

Me parece que la correlación es mínima y, de hecho, una regulación digital que históricamente no ha ponderado correctamente los costes frente a los beneficios generados²³, acarrea, entre otros, unos mayores costes de producción empresarial, lo que perjudica a las pequeñas y medianas empresas europeas en un mercado mundial, es decir, puede dificultar su escalabilidad internacional frente a las grandes multinacionales.

Aunque en el “Reglamento de Inteligencia Artificial de la UE” se incluyen algunas medidas para el apoyo a la innovación de Pymes y empresas emergentes, es probable que entremos en un bucle en el que la sobrerregulación generalizada dificulte esa innovación y continúe socavando nuestra pérdida de competitividad digital, lo que se traduciría en una paulatina pérdida de peso económico e influyente de la UE en el mundo que, a su vez, hace disminuir las posibilidades de empleos de alto valor añadido para los jóvenes, caldo de cultivo para un descontento social que cuestiona el formato de la

23 Moreno, L., y Pedreño, A. op. cit.

Unidad Europea como demostraron los resultados las elecciones europeas del pasado mes de junio.

Ante este escenario, la cuestión central es la siguiente: si desde Europa no somos capaces de situarnos en la vanguardia tecnológica, ¿podremos mantener nuestros valores? Recordemos que gracias a esa innovación se están protegiendo los valores (y fronteras) occidentales, como pone de manifiesto la aplicación de drones de Starlink, propiedad de Elon Musk, para contener el avance ruso en Ucrania²⁴.

Creo que en este ejemplo reside una de las claves para preservar nuestros valores y avanzar en innovación: citaba en este artículo la disputa por la hegemonía mundial entre Estados Unidos y China, donde la segunda poco a poco puede imponerse. De este modo, parece indispensable para Europa, y también para Estados Unidos, una fuerte alianza entre ambas puesto que compartimos valores, una alianza que defina un marco regulatorio común que aquilate la preservación de esos valores y estimule la innovación europea como lo hace la estadounidense²⁵. En otros términos, propiciar un efecto Bruselas-Washington podría facilitar que los valores occidentales penetrasen en mayor medida en terceros países auspiciando un impulso a la innovación europea.

24 <https://cnnespanol.cnn.com/2024/03/26/ucrania-starlink-guerra-drones-rusia-uso-elon-musk-trax/#0>

25 Resulta oportuno destacar que, como recoge Jans Ulrich Gumbrecht en su obra "El espíritu del mundo en Silicon Valley", en Universidades como Stanford, los mejores estudiantes escogen de forma voluntaria filosofía -con un enfoque occidental en cuyas clases "leen clásico tras clásico"- o literatura comparada para cultivar un enfoque de "pensamiento arriesgado" entendido como una vocación de las humanidades sin el cual no puede haber cambios en las instituciones y en los sistemas. Así, armonizar programas formativos universitarios donde se compartiesen nuestros valores occidentales podría ser un buen punto de partida para esta gran alianza que se propone.

La necesidad del pensamiento humanista para afrontar este cambio disruptivo también es descrita por Kissinger, Schmidt y Huttenlocher en su publicación "La era de la inteligencia artificial": "Pero mientras crece el número de individuos capaces de crear IA, las filas de quienes contemplan esta tecnología para la humanidad -sociales, legales, filosóficas, espirituales, morales- siguen siendo peligrosamente escasas."

3. EL PAPEL QUE PUEDE JUGAR ESPAÑA.

Si Europa ha perdido influencia, en el caso de España hemos pasado de generar el 2,14% de la riqueza mundial en 2010 al 1,5% en 2023, lo que indica la débil competitividad de la economía española tras la crisis inmobiliaria.

Con todo, tenemos muchos puntos fuertes para ser optimistas, por ejemplo, en materia energética²⁶, o en la propia IA, donde podemos convertirnos en la gran referencia europea. Veamos algunas de esas fortalezas:

- Nuestras ciudades de tamaño medio (desde una escala de tamaño internacional): actualmente, son las ciudades medias frente a las grandes megalópolis y desarrollos de baja densidad residencial las que están concentrando la innovación tecnológica. Así, en California y durante este siglo, la inversión está basculando hacia San Francisco frente a Silicon Valley, pues si en 2001 la inversión de capital riesgo en San Francisco era de un 10% frente al 90% en Silicon Valley²⁷, desde hace una década ha pasado a ser del orden de un 50%²⁸.

En el caso europeo, las ciudades de tamaño medio donde brota la innovación podrían serían Lisboa o Dublín, además de algunas españolas que luego mencionaremos.

- Ese ecosistema californiano se complementa con ciudades pequeñas como Berkeley, donde, en pleno centro y en su edificio más emblemático se ubica su principal centro de emprendimiento. Espacial y

26 Ligado a nuestro potencial para el despliegue de energías renovables, España puede convertirse en el gran líder europeo en producción y distribución (Magreb-Centroeuropa) de hidrógeno verde -clave para descarbonizar industria y movilidad pesadas-, como se desprende de la primera subasta de hidrógeno verde publicada por el Banco Europeo de Hidrógeno en la que, de los 132 proyectos presentados, los menores costes procedían de proyectos ubicados en la Península Ibérica. De esta forma, de los 7 proyectos escogidos, 3 han sido españoles y 2 portugueses. https://climate.ec.europa.eu/eu-action/eu-funding-climate-action/innovation-fund/competitive-bidding_en?prefLang=es

27 Necesario anteriormente por los amplios espacios que necesitaban todos los equipos de hardware y la falta y carestía del suelo en San Francisco.

28 <https://siliconvalleyindicators.org/data/economy/innovation-entrepreneurship/private-equity/venture-capital-investment/>

morfológicamente, el tamaño de Berkeley, ubicación de su universidad y conexión con San Francisco recuerda en buena medida al caso de San Vicente, Universidad de Alicante y su conexión con el Área Funcional Alicante-Elche.

Efectivamente, el potencial de las ciudades pequeñas como estimulantes de la creatividad, clave para el desarrollo de la IA y tecnológico en general, lo encontramos, haciendo de nuevo un repaso a los orígenes europeos, en la Atenas de Pericles, la Florencia del Renacimiento o la Jena del Romanticismo, todas ellas ciudades de unos 30.000-50.000 habitantes²⁹, compactas y muy bulliciosas como histórica y actualmente son las ciudades españolas.

- Pero también podemos competir a nivel de grandes Regiones Funcionales Urbanas donde se aglutinan los grandes capitales y la toma de decisiones: en 2030 dispondremos de una red de alta velocidad, la segunda red del mundo en términos de longitud y con una oferta de servicios de bajo coste plenamente asentada, que va a permitir situar a cualquier urbe a menos de tres horas del resto del territorio peninsular (más de 20 millones de habitantes), tiempo crítico para que se produzcan viajes de ida y vuelta en el mismo día, lo que favorecerá la integración de mercados. Este aspecto es especialmente relevante en el caso del Corredor Mediterráneo pues, tras muchas décadas de conexiones precarias, por fin se contará con una oferta de transporte público competitiva y de muy alta calidad que hará posible consolidar, por ejemplo, el proyecto “1.070 km Hub”³⁰.

29 Racionero, L. (1990). *Florencia de los Médicis*. Planeta.

30 En esencia, en esta iniciativa se pone de manifiesto “el potencial del Arco Mediterráneo Sur español para convertirse en un líder global en innovación y transformación digital. Destaca el rol del 1.070 Km Hub como catalizador de esta transformación y presenta un análisis exhaustivo sobre las oportunidades, retos y propuestas para fortalecer el ecosistema digital de la región.” (Pedreño, A., Torres, A., Mora, T., y Pérez, E. del M. (2023). 1.070 Km Hub: La mayor alianza del ecosistema de innovación de España. Publicación independiente).

- Y ahora, si pensamos en los sectores en los que somos punteros internacionalmente, sin duda, debemos citar el turismo, en el que España es la segunda potencia mundial por número de visitantes extranjeros anuales, así como el “residencialismo”, que cada vez anota más personas de otros países que vienen a vivir a España en edad laboral favorecidos por el teletrabajo, siendo su máximo exponente los nómadas digitales. Es decir, formamos parte de un país con gran experiencia hospitalaria y que, además, se está posicionado durante los últimos años como muy tolerante a escala global, por lo que partimos con una gran ventaja competitiva para atraer el talento mundial.

Si a todo ello sumamos, una Universidad que enfatice su función de transferencia e incorpore en sus planes de estudios de la rama técnica las humanidades, una regulación que aquilate la preservación de nuestros valores y el estímulo a la innovación, y una intensa campaña de concienciación social sobre lo mucho que está en juego y nuestras fortalezas para liderar esta transformación, estaríamos en condiciones de propiciar su implementación mediante una colaboración público-privada que nos permita afrontar el futuro con seguridad e ilusión. 

BIBLIOGRAFÍA:

1. Allison, G. (2018). *Destined for War: Can America and China Escape Thucydides' Trap?* Scribe.
2. Banco Mundial, GDP (constant 2017 US\$), 2017, <https://datos.bancomundial.org/indicador/NY.GDP.MKTP.PP.KD?end=2023>.
3. Banco Mundial, GDP (current US\$), 2023, <https://datos.bancomundial.org/indicador/NY.GDP.MKTP.CD?end=2023>.
4. Bradford, A. (2020). *The Brussels Effect: How the European Union Rules the World*. Oxford University Press.
5. Brzezinski, Z. (1998). *El gran tablero mundial*. Ediciones Paidós.
6. Castellón, J. J. (2023). *Confucio: maestro de moral, filósofo de ética*. Isidorianum.
7. Chomón, J. M. (2023). *La era de las tierras raras: La cruzada geopolítica por los metales estratégicos*. Tecnos.

8. CNN. (2024, marzo 26). *Ucrania enfrenta problemas con Starlink mientras Rusia aumenta su uso de drones*. CNN en Español. <https://cnnespanol.cnn.com/2024/03/26/ucrania-starlink-guerra-drones-rusia-uso-elon-musk-trax/#0>
9. Comisión Europea. (n.d.). Fondo de Innovación: Licitación competitiva. https://climate.ec.europa.eu/eu-action/eu-funding-climate-action/innovation-fund/competitive-bidding_en?prefLang=es
10. Council of the European Union. (2024). Artificial Intelligence Act: Council Gives Final Green Light to the First Worldwide Rules on AI. <https://www.consilium.europa.eu/es/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/>.
11. Diario Oficial de la Unión Europea. (2024). Reglamento (UE) 2024/1689. https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=OJ:L_202401689.
12. Engineering News-Record (ENR). (2023). 2023 Top 250 International Contractors Preview. <https://www.enr.com/toplists/2023-Top-250-International-Contractors-Preview>.
13. European Commission. (2023). European Chips Act. <https://digital-strategy.ec.europa.eu/es/policies/european-chips-act>.
14. European Commission. (n.d.). Science Rocks: Rare Earths Technology We Can't Live Without. <https://projects.research-and-innovation.ec.europa.eu/en/horizon-magazine/science-rocks-rare-earth-technology-cant-live-without>.
15. European Union. (2020). Publication on the European Union's Research and Innovation Projects. <https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>.
16. Floris, G. (2022). *Panorama de la historia universal del derecho*. Marcial Pons.
17. García, F. (2023). *Érase una vez Europa: Senderos de justicia, tolerancia y libertad*. Espasa.
18. Gobierno de la República Popular China. (2017). Política sobre el desarrollo de la industria de servicios modernos. https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm.

- 19 . Gumbrecht, H. U. (2020). *El espíritu del mundo en Silicon Valley*. Deusto.
- 20 . ICEX. (2025). Plan 2025: China. https://www.observatoriorli.com/docs/CHINA/PLAN_2025_China.pdf.
- 21 . Kai-Fu Lee (2020). *Superpotencias de la inteligencia artificial: China, Silicon Valley y el nuevo orden mundial*. Deusto.
- 22 . Kissinger, H. A., Schmidt, E., & Huttenlocher, D. (2023). *La era de la inteligencia artificial y nuestro futuro humano*. Anaya.
- 23 . Lee, K.-F. (2020). *Superpotencias de la inteligencia artificial: China, Silicon Valley y el nuevo orden mundial*. Planeta.
- 24 . Miller, C. (2023). *La guerra de los chips: La gran lucha por el dominio mundial*. Península.
- 25 . Moreno, L., y Pedreño, A. (2020). *Europa frente a EE.UU. y China. Prevenir el declive en la era de la inteligencia artificial*. Publicación independiente.
- 26 . Pedreño, A., Torres, A., Mora, T., y Pérez, E. del M. (2023). *1.070 Km Hub: La mayor alianza del ecosistema de innovación de España*. Publicación independiente.
- 27 . PwC. (2024). Global Top 100 Companies – by Market Capitalisation. <https://www.pwc.co.uk/audit/assets/pdf/global-100/global-top-100-companies-2024.pdf>.
- 28 . Racionero, L. (1990). *Florecia de los Médicis*. Planeta.
- 29 . Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo. (2016). Reglamento general de protección de datos. <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A32016R0679>.
- 30 . Silicon Valley Indicators. (n.d.). Venture Capital Investment. <https://siliconvalleyindicators.org/data/economy/innovation-entrepreneurship/private-equity/venture-capital-investment/>.

ÉTICA DE LOS SISTEMAS DE IA: CONSTRUYENDO UNA IA TRANSPARENTE, EQUITATIVA Y FIABLE

💬 Juan Pablo Peñarrubia Carrión

La vertiente ética es una de las cuestiones más trascendentes del impacto de las tecnologías informáticas. No obstante, en las pasadas décadas, esta dimensión quedó frecuentemente como asignatura pendiente ante el deslumbramiento de las continuas innovaciones informáticas y su potencialidad e incesante demanda en forma de productos o servicios en todas las esferas de actividad humana. De hecho, la principal causa de esta carencia ha estado en el pretendido paradigma de la autorregulación de los productos, servicios y actividades informáticas, promovido durante décadas por los grandes gigantes tecnológicos. Permitida la selva de no regulación legal, la vertiente aún más compleja del impacto ético quedaba automáticamente excluida de las prioridades en el desarrollo y uso de las tecnologías informáticas. Este escenario comenzó a cambiar alimentado con dos factores clave (Peñarrubia, 2024):

- En primer lugar sonados escándalos mediáticos que ponían en primer plano la vertiente ética de los productos y servicios informáticos. Por ejemplo, el caso Volkswagen de manipulación del software de control de emisiones de sus motores; el caso Facebook-Cambridge Analytica de recopilación y uso masivo de datos de ciudadanos con fines de propaganda política; y el caso Boeing 737 MAX, tras dos accidentes en octubre de 2018 y marzo de 2019 y la muerte de 346 personas.
- Y por otro lado, la constatación, sin tapujos, de los usos malintencionados de las tecnologías informáticas para influir en procesos sociales

(elecciones presidenciales de EEUU¹, referéndum del BREXIT² o en España la acción rusa con el independentismo catalán³...), promover desinformación mediante noticias falsas y procesos de viralización artificial, agitación social... En definitiva, usos dañinos para la desestabilización y manipulación social, especialmente en las sociedades democráticas abiertas. Frecuentemente fruto de actuaciones desde estados extranjeros totalitarios, con técnicas que actualmente se consideran tácticas de ciberguerra.

En este escenario, las últimas innovaciones en IA han hecho de catalizador de las inquietudes pendientes de resolver en materia de impacto ético de la informática. La IA ha cristalizado la "espoleta social" para exigir medidas de regulación y control de las tecnologías informáticas tras décadas de preocupación creciente en relación a los productos y servicios informáticos que han construido la sociedad digital: indefensión, opacidad, desresponsabilización, cosificación, amenaza de los cimientos de la sociedad... (Peñarrubia, 2023) Asistimos a un cambio de visión para poner en primer término la importancia del impacto ético de las tecnologías informáticas, liquidando el paradigma de la autorregulación y promoviendo el paradigma de la regulación necesaria y la ética (Floridi, 2021). En el aspecto de la regulación necesaria por primera vez se aborda una regulación con dos vertientes: normativa legal y normativa de estandarización, con la ambición de conciliar la garantía de los derechos fundamentales y algunos cimientos sistémicos como la seguridad y la democracia real, con la promoción de la innovación. En el ámbito europeo se trata de una variante de la denominada regulación inteligente / *smart regulation*. En

1 EE. UU. asegura que Rusia e Irán intentaron manipular el resultado de las presidenciales de 2020. Diario El País. 13 de marzo de 2021. Disponible en: <https://elpais.com/internacional/2021-03-16/ee-uu-asegura-que-rusia-eiran-intentaron-manipular-el-resultado-de-las-presidenciales-de-2020.html>

2 Reino Unido admite que la «influencia rusa» en sus procesos democráticos «se ha convertido en la nueva normalidad» Diario El Mundo. 21 de julio de 2020. Disponible en: <https://www.elmundo.es/internacional/2020/07/21/5f16c53221efa094198b45ba.html>

3 Los 4.800 bots que jalearon el 'procés'. Diario El País. 4 de diciembre 2017. Disponible en: https://elpais.com/politica/2017/12/04/actualidad/1512389091_690459.html

este nuevo contexto la dimensión ética de la IA, y por extensión de cualquier tecnología informática, se pone en primer término, por un lado impregnando los desarrollos regulatorios tanto a nivel legal como de estandarización, y en segundo lugar promoviendo medidas para una ética aplicada de la IA que materialice valores clave como la equidad, la fiabilidad o la transparencia.

IMPOSIBILIDAD DE UNA ÚNICA APROXIMACIÓN ÉTICA:

Pero la visión ética (y también regulatoria) de la IA no es uniforme a nivel global. La fundamentación ética depende del grupo social concernido: cultura, tradición, ideología... Lo cual lleva a una fundamentación ética susceptible de diferentes variantes, como se está reflejando en las diferentes aproximaciones regulatorias y éticas de la IA a nivel global: EU, USA, China... (Domingo y Peñarrubia, 2024a; 2024b) Para ilustrarlo, y en contraste a la visión de la UE, que detallaremos al final de este epígrafe, veamos un par de pinceladas distintivas de la aproximación China. Por ejemplo, la aplicación Ernie, un robot conversacional o *chatbot* basado en IA generativa, con un enfoque similar a ChatGPT, es reticente a responder a preguntas sobre temas como la COVID-19, o sobre líderes políticos chinos⁴. En el contexto chino, medidas como esta están alineadas con la acción habitual del gobierno por lo que no son percibidas por la sociedad china como algo nuevo significativo, sino como parte de los modos de hacer de la sociedad china. Por el contrario, la visión de los medios de comunicación europeos remarca la incidencia en el derecho a la información, hablando abiertamente de censura. En abril de 2023 el gobierno chino anunciaba "medidas administrativas para servicios de inteligencia artificial generativa", estableciendo que "El contenido generado por la inteligencia artificial generativa debe encarnar los valores socialistas fundamentales y no debe contener ningún contenido que subvierta el poder estatal, abogue por el derrocamiento del sistema socialista, incite a la división del país o socave la unidad nacional", prohibiéndose de hecho todo contenido "que pueda

4 Ernie, la inteligencia artificial china entrenada con censura informativa. <https://www.la-vanguardia.com/tecnologia/actualidad/20230525/8990969/china-censura-propio-chatbot-inteligencia-artificial-pmv.html>)

perturbar el orden económico y social”⁵. En julio de 2024 se informaba del modo de materializar este control desde la Administración del Ciberespacio de China⁶. De nuevo, en el contexto chino este tipo de acciones forman parte de lo cotidiano por lo que no son percibidas por la sociedad china como un impacto ético relevante. Sin embargo, la visión de los medios de comunicación occidentales subraya la ideologización e incidencia en las libertades básicas y en el derecho a la información.

Dada la imposibilidad de una única fundamentación ética es clave el impulso de una ética aplicada ligada a referentes comunes a escala mundial. Es decir, que aun pudiendo existir diferencias, se tome en consideración un cierto marco común. Este marco global común, en primera instancia, estaría dado por la ética ligada a la protección de los derechos contenidos en La Declaración Universal de los Derechos Humanos⁷. A nivel europeo el referente común es la Carta de los Derechos Fundamentales de la Unión Europea, que va más allá de la declaración de derechos humanos recogiendo derechos civiles, políticos y sociales de los ciudadanos de la Unión Europea⁸.

Conviene indicar, en este contexto, que cuando se habla de la Declaración Universal de los Derechos Humanos estamos en un terreno con al menos dos dimensiones: En primer lugar, la vertiente ética pura ligada al respeto a los derechos contenidos en la Carta, y por otro lado, la vertiente jurídica derivada del hecho de suscribir un estado un tratado internacional, en este caso la

5 China regula que sus modelos de inteligencia artificial tengan “valores socialistas”. Diario La Razón. 12 de abril de 2023. Disponible en: https://www.larazon.es/tecnologia/china-regula-que-sus-modelos-inteligencia-artificial-tengan-valores-socialistas_202304126435c98d1b5f5b00013feg95.html

6 ¿Una IA comunista? China usa a censores para comprobar si sus IA “encarnan valores socialistas”. Diario Euronews. 21 de julio de 2024. Disponible en: <https://es.euronews.com/next/2024/07/21/ai-comunista-china-usa-a-censores-para-comprobar-si-los-modelos-de-ia-encarnan-los-valores>

7 Declaración Universal de Derechos Humanos. Disponible en <https://www.un.org/es/universal-declaration-human-rights/>

8 Carta de los Derechos Fundamentales de la Unión Europea. Disponible en: <https://eur-lex.europa.eu/ES/legal-content/summary/charter-of-fundamental-rights-of-the-european-union.html>

Declaración Universal de los Derechos Humanos. Ello conlleva una obligación, pero las cartas de derechos no concretan, ni detallan los mecanismos para su implementación y tampoco las consecuencias de su incumplimiento. Por lo que, si bien se trata de un gran avance a nivel internacional, es necesario un desarrollo para su repercusión práctica sistémica en el día a día de los estados y sociedades. En el caso que nos ocupa para la traslación de estos derechos a los sistemas de IA. Y lo que es igual de importante, los mecanismos de control y las consecuencias de los incumplimientos. Este terreno de la aplicación práctica es el gran reto: tanto a nivel de la regulación legal como de la ética aplicada en general, que se despliega más allá de las obligaciones legales estrictamente hablando, y en todos los ámbitos sociales: empresas, administración pública, profesionales, ciudadanía... La ética sin aplicación práctica es mero academicismo, o lo que es peor en el ámbito tecnológico: desde brindis al sol a demagogia interesada.

LOS ROLES DE IA:

A la hora de evaluar y gestionar el impacto ético del uso de la IA existen grandes diferencias dependiendo del rol de la organización o individuo en cuestión en relación al sistema de IA. Para comprenderlo fácilmente basta pensar en una organización desarrolladora de un sistema de IA versus una organización meramente usuaria del mismo sistema de IA, o versus un ciudadano afectado por las decisiones de un sistema de IA... Por lo tanto, a la hora de evaluar y gestionar cuestiones éticas de un sistema de IA una dimensión muy importante a tener en cuenta serán los diferentes roles de IA en relación a dicho sistema. No existe una lista exhaustiva de los roles de IA pero sí una taxonomía de referencia que incluye figuras como: proveedor de IA, productor de IA, cliente de IA, usuario de IA, colaborador de IA, integrador de sistemas de IA, evaluador de IA, auditor de IA, sujeto de IA, autoridad relevante de IA, etc. (ISO, 2022a)

ÉTICA DE LOS SISTEMAS DE IA EN LA UE

Un elemento esencial y distintivo de la visión europea de la IA es la consecución de un uso ético de los sistemas de IA. De hecho este espíritu se recoge explícitamente en el Reglamento europeo de Inteligencia Artificial (en adelante

Reglamento de IA): “el presente Reglamento respalda el objetivo de promover el enfoque europeo de la IA centrado en el ser humano y de ser un líder mundial en el desarrollo de IA segura, digna de confianza y ética, como indicó el Consejo Europeo, y garantiza la protección de los principios éticos, como solicitó específicamente el Parlamento Europeo” (Comisión Europea, 2024b)

Con dicho objetivo a nivel de regulación legal y gobernanza de la IA, la Comisión Europea promovió un estudio sobre cuales han de ser las directrices éticas para una IA confiable y que respete la Carta de los Derechos Fundamentales de la Unión Europea. El informe de 2019 “Directrices éticas para una IA fiable” del Grupo independiente de expertos de alto nivel sobre inteligencia artificial de la Comisión Europea, establece un conjunto no exhaustivo de requisitos para una IA fiable⁹. (Figura 1) Todos los requisitos se consideran de idéntica importancia, se apoyan entre sí y deberían cumplirse y evaluarse a lo largo de todo el ciclo de vida de un sistema de IA. (AI-HLEG, 2019) Estos mismos requisitos son asumidos como los “principios éticos” que guían del Reglamento de IA: “Los siete principios son: acción y supervisión humanas; solidez técnica y seguridad; gestión de la privacidad y de los datos; transparencia; diversidad, no discriminación y equidad; bienestar social y ambiental, y rendición de cuentas. Sin perjuicio de los requisitos jurídicamente vinculantes del presente Reglamento y de cualquier otro acto aplicable del Derecho de la Unión, esas directrices contribuyen al diseño de una IA coherente, fiable y centrada en el ser humano, en consonancia con la Carta y con los valores en los que se fundamenta la Unión.” .../...“Se anima a todas las partes interesadas, incluidos la industria, el mundo académico, la sociedad civil y las organizaciones de normalización, a que tengan en cuenta, según proceda, los principios éticos para el desarrollo de normas y mejores prácticas voluntarias.” (Comisión Europea, 2024b)

Nótese que en la visión europea, estos principios no solo vertebran la regulación legal contenida en el Reglamento de IA, sino que además se consideran esenciales para la confiabilidad de la IA en el desarrollo de “normas y mejores

9 Estos requisitos tienen su origen en la Carta de los Derechos Fundamentales de la Unión Europea

prácticas voluntarias". Ello es muy relevante para el ámbito ético en sentido amplio que, como es sabido, abarca más allá de la estricta obligación legal. Y también para el ámbito de las organizaciones de normalización, como productoras de estándares sectoriales que, incluso siendo de uso voluntario, constituyen una potente herramienta de gobernanza de la industria ligada a cualquier tecnología. Sobre este aspecto, se abundará en el epígrafe sobre gobernanza de la IA.

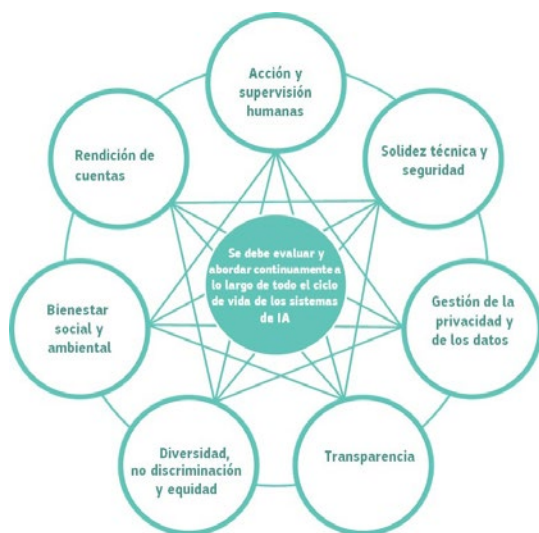


Figura 1. Siete requisitos esenciales y sus interrelaciones para una IA confiable. Fuente (AI-HLEG, 2019)

Adicionalmente a este profundo fundamento ético de la regulación legal europea materializada en el Reglamento de IA, se están impulsando desarrollos normativos para materializar una ética aplicada de los sistemas de IA. En definitiva, para llevar la gestión de cuestiones ética de IA al día a día de las empresas, las administraciones, los consumidores, los ciudadanos en general y la sociedad en su conjunto. Se trata de normas de estandarización que se están realizando a nivel europeo en el Comité Conjunto de Estandarización CEN/CLC/JTC 21 - Artificial Intelligence, cuya cometido en la visión europea de la IA se detalla en el epígrafe sobre gobernanza de la IA. No obstante, a continuación se recogen brevemente las actuales iniciativas en curso con especial importancia o incidencia en el ámbito de la ética de la IA a nivel europeo¹⁰:

¹⁰ Extracto de los títulos y alcances de algunas iniciativas de estandarización a fecha de redacción de esta publicación.

- Marco de fiabilidad de la IA (*AI trustworthiness framework*): Dirigido principalmente a las organizaciones que comercializan o ponen en servicio sistemas de IA. Proporciona un marco para la fiabilidad de los sistemas de IA que contiene terminología, conceptos, requisitos horizontales de alto nivel, orientación y un método para contextualizarlos específicamente a partes interesadas, dominios o aplicaciones. Los requisitos horizontales de alto nivel abordan los aspectos fundamentales y la caracterización de la fiabilidad de los sistemas de IA. Previsión de disponibilidad en 2025
- Requisitos de competencia para profesionales de ética de la IA (*Competence requirements for AI ethicists professionals*): Proporciona un marco sistematizado para las competencias del especialista en ética de la IA, categorizando conocimientos, habilidades y actitudes relacionados con las actividades y tareas específicas de este perfil. Identifica los requisitos y recomendaciones necesarios para que las personas puedan ejercer eficazmente como especialistas en ética de la IA. Previsión de disponibilidad en 2025.
- Directrices sobre herramientas para la gestión de cuestiones éticas durante el ciclo de vida de los sistemas de IA (*Guidelines on tools for handling ethical issues in AI system life cycle*)¹¹: Proporciona especificaciones y directrices para un conjunto de herramientas (instrumentos, actividades, procedimientos) para la gestión práctica de preocupaciones sociales y éticas a lo largo del ciclo de vida de los sistemas de IA: desde la aparición de una nueva cuestión hasta la correspondiente toma de decisiones. Se aporta una lista de herramientas junto con un conjunto mínimo de requisitos y orientaciones que permiten a las organizaciones ponerlas en práctica: se trata de evitar que cada organización deba reinventar la rueda para gestionar cuestiones de ética de la IA. Como principio general, las herramientas proporcionadas son independientes de una fundamentación ética concreta. Aplicable a todas las organizaciones, con cualquier rol de IA, incluidos clientes y usuarios. No obstante, el objetivo principal es ayudar

11 España lidera dos iniciativas de estandarización de la Inteligencia Artificial en Europa. Asociación Española de Normalización (UNE). Disponible en: <https://www.une.org/la-asociacion/sala-de-informacion-une/noticias/espana-lidera-dos-iniciativas-de-estandarizacion-de-la-inteligencia-artificial-en-europa>

a los productores y proveedores de sistemas de IA en la gestión práctica de cuestiones sobre preocupaciones sociales y consideraciones éticas durante el ciclo de vida de los sistemas de IA. Previsión de disponibilidad en 2026. Esta iniciativa es especialmente trascendente para las empresas pues la existencia de herramientas prácticas hace posible que los profesionales de las empresas proveedoras de IA puedan incorporar la resolución de cuestiones éticas como un tipo más de requisitos prácticos en el ciclo de vida de los sistemas de IA. De este modo, las empresas proveedoras de IA podrán adoptar una ética corporativa aplicada a los sistemas de IA de modo viable y sostenible (Domingo y Peñarrubia, 2024b).

- Orientaciones para mejorar las capacidades de las organizaciones sobre ética y preocupaciones sociales asociadas a la IA (Guidance for upskilling organisations on AI ethics and social concerns): Proporcionará orientación para la mejora de las capacidades de las organizaciones en materia de ética y preocupaciones sociales asociadas a la IA. El objetivo final es permitir a las organizaciones mejorar sus capacidades básicas y su cultura en relación con la ética de la IA, tanto a nivel individual (para los empleados en cualquier puesto) como a nivel organizacional (incluidos los equipos de desarrollo de IA, departamentos, directivos, etc.). En particular, para desarrollar programas de formación continua. Aplicable a todas las organizaciones, con cualquier rol de IA. Previsión de disponibilidad en 2026.

Adicionalmente, es conveniente mencionar otros desarrollos en curso de normas de estandarización con especial importancia o incidencia en el ámbito de la ética de la IA a nivel mundial global, en el comité ISO/IEC JTC 1/SC 42 Artificial intelligence. A veces se trata de proyectos coordinados de modo paralelo (*parallel developments* CEN/CLC – ISO/IEC), es decir, que dan lugar a normas tanto europeas como globales¹². Y en otros casos se trata de desarro-

12 En el momento de elaboración de esta publicación se está analizando la posibilidad de desarrollo paralelo CEN/CLC – ISO/IEC de las iniciativas de Directrices sobre herramientas para la gestión de cuestiones éticas durante el ciclo de vida de los sistemas de IA y Orientaciones para mejorar las capacidades de las organizaciones sobre ética y preocupaciones

llos globales que pueden ser adoptados a nivel europeo, ya sea literalmente o con incorporación de especificidades europeas. Esta última posibilidad es muy interesante para promover normas de estandarización globales y al mismo tiempo coherentes con el modelo europeo de regulación legal y de fundamentación ética.

- ISO/IEC TR 24368 – Tecnología de la Información – Inteligencia Artificial – Perspectiva general de las preocupaciones éticas y sociales (*ISO/IEC TR 24368:2022 Information technology – Artificial intelligence – Overview of ethical and societal concerns*) (ISO, 2022c): Visión general de alto nivel de los problemas éticos y sociales asociados la IA. Además: proporciona información en relación con los principios, procesos y métodos en este ámbito; está destinado a tecnólogos, reguladores, grupos de interés y la sociedad en general; no pretende defender ningún sistema de valores específico. Publicado en 2022.
- ISO/IEC AWI TS 22443 – Tecnología de la Información – Inteligencia Artificial – Orientaciones para abordar las preocupaciones sociales y las consideraciones éticas (*ISO/IEC AWI TS 22443 Information technology – Artificial intelligence – Guidance on addressing societal concerns and ethical considerations*) (ISO, 2023c): Proporciona orientación sobre cómo una organización puede identificar y abordar las preocupaciones sociales y las consideraciones éticas durante el ciclo de vida de los sistemas de IA que potencialmente pueden dañar a las personas y a la sociedad. Aplicable a todo tipo de organizaciones que desarrollen o utilicen sistemas de IA, independientemente de su tamaño, tipo y naturaleza. Previsión de disponibilidad en 2025.

Finalmente, junto a los aspectos éticos ligados directamente a los sistemas de IA existen implicaciones éticas más amplias ligadas a los usos potenciales de la IA en sistemas integrados, especialmente en su interacción

sociales asociadas a la IA. Si la decisión es positiva el resultado será una normalización tanto europea como global.

con personas. Especial mención en este campo es la aplicación de la IA en la robótica, ya se trate de avatares, o robots software en general, o bien de robots físicos. Más allá de los aspectos éticos asociados a las funcionalidades en sí, los individuos pueden percibir una “personalidad” donde solo hay una máquina (material o virtual), lo cual puede derivar en la generación de vínculos emocionales, con inusitadas repercusiones potenciales. Especialmente en entornos virtuales, como los videojuegos, o en ambiciosas iniciativas como el metaverso, son necesarias medidas éticas por diseño que, por ejemplo, permitan conocer en todo momento si una persona está interactuando con otro ser humano o bien con un sistema integrado de IA.

REFERENCIAS BIBLIOGRÁFICAS:

1. AI-HLEG (2019) Ethics guidelines for trustworthy AI. High-Level Expert Group on Artificial Intelligence (AI-HLEG). Versión en español: Directrices éticas para una IA fiable. Grupo independiente de expertos de alto nivel sobre inteligencia artificial. Comisión Europea, Oficina de Publicaciones. 2019. Brussels. ISBN 978-92-76-11994-4. doi:10.2759/14078. Disponible en: <https://data.europa.eu/doi/10.2759/346720>. (Consulta: 3 de septiembre de 2024)
2. Comisión Europea (2024b) Reglamento (UE) 2024/1689 Del Parlamento y del Consejo, de 13 de junio de 2024 por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial) Diario Oficial, L 1689, ELI. Disponible en: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj> (Consulta: 3 de septiembre de 2024)
3. Domingo, Agustín; Peñarrubia, Juan Pablo (2024a) Herramientas y ética profesional para impulsar la ética de la IA en la empresa. Proceedings XXXI Congreso Internacional European Business Ethics Network EBEN-Spain 2024: Business humanism in the digital age: An ethical perspective on artificial intelligence. 3-5 Junio 2024. Cáceres. September, 2024 – Language: Spanish, English – Open Access DOI: 10.3926/eben24 / ISBN:

- 978-84-126475-9-4 Disponible en: <https://www.omniascience.com/books/index.php/proceedings/catalog/book/148>
- 4 . Domingo, Agustín; Peñarrubia, Juan Pablo (2024b) Faros en la niebla: Herramientas de ética aplicada para sistemas de IA. Conferencia en XXV World Congress of Philosophy: "Philosophy across Boundaries" . 1-8 Agosto 2024. Roma. Sitio web del congreso: <https://wcprome2024.com>
 - 5 . Floridi, Luciano (2021). The End of an Era: from Self-Regulation to Hard Law for the Digital Industry. *Philosophy and Technology*, 34(4), 619–622. [https:// doi.org/10.1007/s13347-021-00493-0](https://doi.org/10.1007/s13347-021-00493-0)
 - 6 . ISO (2022a) ISO/IEC 22989:2022 Information technology — Artificial intelligence — Artificial intelligence concepts and terminology. International Organization for Standardization – ISO. Ginebra. Disponible en: [https:// www.iso.org](https://www.iso.org) (Consulta: 8 de septiembre de 2024)
 - 7 . ISO (2022c) ISO/IEC TR 24368:2022 Information technology — Artificial intelligence — Overview of ethical and societal concerns. International Organization for Standardization – ISO. Ginebra. Disponible en: [https:// www.iso.org](https://www.iso.org) (Consulta: 8 de septiembre de 2024)
 - 8 . ISO (2023c) ISO/IEC AWI TS 22443 Information technology — Artificial intelligence — Guidance on addressing societal concerns and ethical considerations. International Organization for Standardization – ISO. Ginebra. Disponible en: <https://www.iso.org> (Consulta: 8 de septiembre de 2024)
 - 9 . Peñarrubia, Juan Pablo (2023) ¿Puede materializarse una ética de la inteligencia artificial, o incluso una ética digital, sin una ética profesional informática consolidada? Conference on XXI International Congress on Ethics and Political Philosophy / Congreso Internacional de Ética y Filosofía política: Ética, Democracia e Inteligencia artificial. 2-4 February 2023. Asociación Española de Ética y Filosofía Política (AEEFP). University Jaume I. Castellón.
 - 10 . Peñarrubia, Juan Pablo (2024) ¿Quién cuida el jardín?: Cuidado humanizador de los cimientos individuales y sociales ante el impacto de las tecnologías digitales *SCIO. Revista de Filosofía*, n.º 26, Julio de 2024, 123-149, ISSN: 1887-9853 DOI: https://doi.org/10.46583/scio_2024.26.1155. 

IMPACTO DE LA IA EN LA SOCIEDAD: EMPLEO, ADMINISTRACIÓN, CIUDADANÍA Y POLÍTICA

💬 Juan Pablo Peñarrubia Carrión

El impacto de la IA en la sociedad afecta a prácticamente todos los sectores de actividad. Nada nuevo en la historia de innovación de las tecnologías informáticas en las últimas décadas. Las diferencias del impacto actual de la IA nacen esencialmente de tres elementos:

- La gestión de problemas o situaciones hasta ahora no resueltos con la suficiente efectividad por la informática clásica (algorítmica). Esta es la esencia de la IA en el conjunto de la ingeniería informática: resolver problemas que o bien no admiten soluciones algorítmicas o bien dichas soluciones no son suficientemente eficientes. Este ha sido el principal ámbito de acción de la IA en las últimas décadas.
- Las posibilidades de la IA generativa. La gran novedad en el panorama de la IA ha sido la efectividad de la IA generativa, con especial mención del hito del lanzamiento de la aplicación ChatGPT hace menos de dos años¹.
- La sinergia de la IA, especialmente la IA generativa, con el resto sistemas y tecnologías informáticas existentes que está originando un aluvión de posibilidades de innovación. En este escenario tiene especial importancia el impulso de la automatización que, si bien no es una característica intrínseca de la IA, es una clara fuente de productividad, por lo que la combinación de la IA con otras tecnologías informáticas está disparando las capacidades de automatización en todos los ámbitos.

1 En concreto desde el 30 de noviembre de 2022 en que OpenAI lanzó la aplicación ChatGPT. Un chatbot de inteligencia artificial generativa especializado en diálogo y generación de contenidos escritos. Las amplitud de temáticas de los contenidos, los buenos resultados en la percepción general de los usuarios, y su uso gratuito resultaron en un crecimiento e intensificación de su utilización a escala global: un millón de usuarios en el primer mes. Cien millones de usuarios al cabo de un año..

La incesante actividad comunicativa y opinativa sobre la IA y las oportunidades y amenazas asociadas está generando una gran preocupación social. De hecho, hay estudios estadísticos que muestran que en España siete de cada diez personas percibe la IA como una amenaza para la desinformación, los ciberataques o la privacidad². Al tratarse de una percepción ciudadana, es seguro que está integrando no solo la pura IA, sino inquietudes en relación a las tecnologías de la información en general o a aspectos negativos de las dinámicas sociales de los últimos tiempos: redes sociales, polarización, videovigilancia... Pero, de cualquier modo, refleja la expectación generalizada ante el impacto de la IA.

A continuación, se aporta una aproximación al impacto de la IA en algunos ámbitos concretos por su especial trascendencia a nivel social general.

EL IMPACTO DE LA IA EN EL EMPLEO:

Según un estudio sobre la evolución del empleo a nivel mundial, el 50% de empresas cree que la IA generará un crecimiento del empleo, mientras que el 25% cree que causará una pérdida de puestos de trabajo (WEF, 2023).

En España, se estima que el uso de la IA durante la próxima década afectará a los empleos existentes, con previsiones de un 9,8% de ellos en riesgo de ser automatizados, un 15,9% que podrían beneficiarse de la IA incrementando su productividad, mientras que el resto de los empleos actuales, prácticamente 3 de cada 4, no se verían afectados significativamente por la IA durante la próxima década. Por otro lado, se estima que la expansión de la IA a nivel empresarial creará nuevas oportunidades y empleos, con una creación de 1,61 millones de empleos nuevos para la próxima década en España. Aunque se prevé un efecto neto de la IA en el empleo ligeramente negativo, con la pérdida potencial de unos 400.000 puestos de trabajo en los próximos 10 años (Randstad, 2024).

Otros estudios estiman que en España: hay 11,1 millones de puestos de trabajo que estarían más expuestos a la IA, por ejemplo la práctica totalidad

2 Estudio Ipsos de noviembre de 2023 disponible en <https://www.ipsos.com/es-es/hali-fax-report-2023-ai>

de los profesionales de servicios financieros, programación y servicios de tecnologías de la información, servicios jurídicos y contables. Por comunidades autónomas la Comunidad de Madrid, Cataluña, País Vasco, Navarra y Cantabria tendrían un porcentaje de ocupaciones expuestas a la IA mayor que la media nacional. (AFI, 2024)

Ciertamente la IA, como lo hicieron antes otras tecnologías informáticas, tendrá un impacto en el empleo. La lección aprendida es que para gestionar sin sobresaltos la transición, tanto a nivel social como empresarial, no hay que ser ni alarmista ni negacionista. Ya hemos vivido importantes transformaciones del trabajo por el impacto de la informática. Más allá de las diferentes oleadas de informatización progresiva que han confluído en la denominada transformación digital hay que recordar que no hace tanto tiempo las empresas funcionaban con documentos en papel a base de máquinas de escribir y papel carbón... Y claro está, legiones de auxiliares administrativos. Todo cambió con la revolución ofimática del ordenador personal y los programas de proceso de textos. Estos avances informáticos eran disruptivos para la sociedad y la empresa de la época, y por supuesto tuvieron alto impacto en el empleo. Pero ¿hay alguna duda sobre lo positivo de gestionar con procesador de textos frente a la máquina de escribir y el papel carbón? La clave de la transición siempre ha sido la flexibilidad de las personas y la capacidad de reciclarse aprovechando la innovación para aumentar la productividad y a la postre la empleabilidad. Del mismo modo, la clave sobre el impacto de la IA en el empleo está en la actitud personal flexible para aprovechar las capacidades de la IA. Para ilustrarlo con ejemplos: Si un abogado puede tener un primer borrador de un recurso en un minuto gracias a la IA en lugar de emplear dos horas en hacerlo a mano, podrá utilizar ese tiempo ahorrado en mejorar la calidad del resultado final, que es lo fundamental. Lo mismo es aplicable a un periodista, o a un programador informático. La cuestión clave será la actitud personal para reciclarse y, claro está, el nuevo perfil de puesto de trabajo resultante en ocupaciones existentes. Y también la aparición de nuevas ocupaciones a tenor de las nuevas posibilidades. Ya hemos asistido a escenarios de impacto sistémico de la informática en el empleo. Y asistiremos a otros en el futuro ¿O alguien cree que la IA será la última innovación informática de alto impacto?

EL IMPACTO DE LA IA EN LA ADMINISTRACIÓN PÚBLICA:

Las oportunidades de aprovechamiento de la IA en el sector público son enormes. Destacando el interés en su aplicación para³:

- Asistentes inteligentes: tanto para la resolución de consultas como la orientación de la ciudadanía en sus necesidades ante la administración. Chatbots de provisión de información, asistentes de orientación a la ciudadanía en sus trámites, asistentes virtuales de atención turística...
- Detección de fraudes y corrupción.
- Ayuda a la decisión en la gestión pública a todos los niveles.
- Mejora de la interoperabilidad: Y con ello impulso de una ventanilla única real en la relación con las administraciones.
- Gestión municipal y urbana: planificación urbana, gestión de residuos, gestión de la movilidad urbana, gestión de la calidad del aire, gestión del ruido...
- Estimación de evolución de ingresos por impuestos, tasas o actividades económicas.
- Automatización de procesos o tareas.
- Detección temprana de enfermedades.
- Predicción de epidemias.
- Refuerzo de la ciberseguridad.
- Optimización del tráfico y transporte público.
- Detección y respuesta a desastres naturales.
- Traducción y servicios multilingües.
- Digitalización automatizada de documentos en papel e integración en sistemas documentales existentes para su puesta en valor.
- Etc. Etc.

Por otro lado, según un reciente estudio sobre el uso de la IA en el sector público a nivel europeo, todos los ámbitos públicos están haciendo ya uso

3 Puede profundizarse en un mayor detalle y casos específicos de aplicación en: (Comisión Europea, 2024a) (AMTEGA, 2023) (Tangi et al, 2022) (Manzoni et al., 2022)

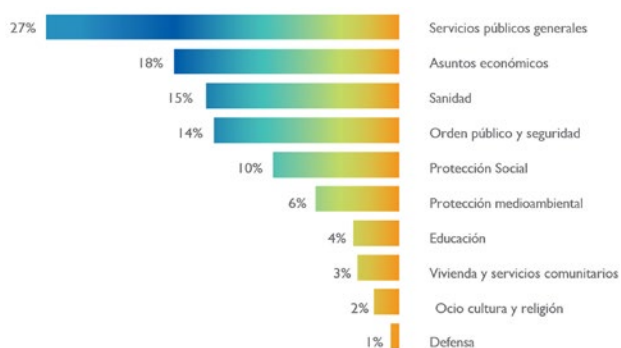


Figura 1. Distribución de casos de uso de IA en el sector público europeo.
Fuente (Comisión Europea, 2024a)

de la AI en las mencionadas aplicaciones o en otros casos de uso de interés, como muestra la Figura 1.

Pero junto a las mencionadas oportunidades, existen potenciales riesgos conocidos (Manzoni et al., 2022): sesgos algorítmicos o de datos, opacidad y complejidad en la explicabilidad y transparencia de las decisiones de los sistemas de IA, incidencia en el empleo, vulneración de la privacidad, polarización social, efectos negativos a nivel medioambiental... A estos riesgos hay que añadir uno especialmente trascendente que subyace en el uso de los sistemas de IA en la administración pública en cuestiones que afectan directamente a personas: la desresponsabilización. Tanto cuando se trate de aplicaciones de ayuda a la decisión, como incluso de automatización de procesos o tareas, debe preservarse la supervisión humana y la delimitación de la responsabilidad de las decisiones ejecutadas. Nótese que la informática ha materializado ya en alto grado la ayuda a la decisión y la automatización en el sector público, pero las actuales oportunidades de automatización gracias a la integración de sistemas y tecnologías con sistemas de IA no debe hacernos caer en la tentación de una huida hacia la desresponsabilización con el pretexto de la eficiencia. No solo porque sería una deshumanización de la administración pública, sino porque introduciría la posibilidad de un sibilino mecanismo de corrupción en el sistema público. A ello habría que añadir los obstáculos potenciales de la IA en el sector público: gestión inadecuada de los datos; acceso insuficiente a grandes volúmenes de datos con la calidad necesaria; interoperabilidad deficiente entre organizaciones; insuficiente dotación humana y/o de recursos en los departamentos de informática; deficiente comprensión de los propios datos en las organizaciones; gobernanza de

datos poco desarrollada: falta de cultura organizativa de la mejora continua; falta de cualificación de personas; competitividad global creciente; legislación y normativa dispersas; falta de confianza; impactos insuficientemente conocidos...

Con este escenario hay que animar a las administraciones públicas a promover la innovación mediante el uso de la IA en procesos en que pueda incrementar la eficiencia y la capacidad de gestión y gobernanza. Tanto por la mejora en sí misma, como también por la capacidad tractora del sector público para promover una tecnología cada día más relevante para la competitividad económica de nuestro territorio en un contexto globalizado. Por supuesto con la debida prudencia, especialmente en lo que respecta a la supervisión humana de los sistemas de IA y la dotación técnica necesaria. Sin los recursos adecuados es imposible la gobernanza tanto de los sistemas que usan IA como de los sistemas informáticos en general en tanto que infraestructura esencial para la función de las administraciones públicas en la sociedad digital.

EL IMPACTO DE LA IA A NIVEL POLÍTICO Y CIUDADANO:

La AI supone una nueva aceleración en los cambios a nivel político y ciudadano a los que venimos asistiendo en las últimas décadas de incesantes desbordamientos derivados de las continuas innovaciones informática: institucionales, legales, en la creación de opinión pública, en la comunicación política, etc.

La incidencia en la gobernanza política, las instituciones y las relaciones con la ciudadanía tiene innumerables vertientes de las que vamos a destacar solo unas pocas por su especial trascendencia sistémica a nivel político, ciudadano y social en general (Peñarrubia 2024):

- En primer lugar, la IA va suponer una nueva vuelta de tuerca en los cambios en los medios de comunicación y en la comunicación política, y con ello en la creación de opinión pública, procesos electorales, etc. En las sociedades abiertas se considera que existe una función social de los medios de comunicación, especialmente en relación con la política y las libertades. Pero hace tiempo que lo digital incide fuertemente en la

dinámica de los medios de comunicación, en particular mediante las redes sociales. La IA permite generar contenidos automatizados, así como la integración con otras tecnologías informáticas para dinamizar o “viralizar” contenidos y amplificar mensajes. La IA propicia incluso generar argumentaciones y relatos para los propios representantes políticos y partidos. Todo ello incide en las dinámicas sociales y en el desarrollo de las campañas electorales. Como se mencionó en el epígrafe sobre ética, están constadas intervenciones con intención de influencia o desestabilización, frecuentemente desde estados extranjeros. Se ponen en práctica técnicas de desinformación y creación de noticias con objeto de influir en la sociedad y generar polarización social. La IA puede amplificar las técnicas actuales de uso de “bots” (robots software), especialmente en redes sociales. En este caso, las capacidades de la IA y la informática en general son también la clave para la solución del problema. Son personas y actores sociales concretos (estados, entidades, etc.) quienes están detrás de estos malos usos. Y solo la IA y la informática en su conjunto pueden alumbrar soluciones viables. En última instancia no solo estamos hablando del derecho a la información, que forma parte de la Carta de Derechos Humanos, sino de un elemento sistémico para el sostenimiento de las sociedades democráticas abiertas en el contexto de la Era de la Información y el Conocimiento.

- Por otro lado, existe también un impacto en la privacidad y la autonomía personal que está conectado con el ámbito político y social en general. La IA incrementa las actuales capacidades de las tecnologías informáticas de incidir en la privacidad de las personas. Esta incidencia no tiene solo una dimensión personal, sino que al afectar a la autonomía personal está también relacionada con la política y la dinámica social. A nadie se le escapa que el “gran hermano” imaginado por Orwell es un juego de niños ante las capacidades de las tecnologías informáticas actuales, especialmente si integran el uso de sistemas de IA. La modernidad de la Constitución Española, permitió recoger en el artículo 18 apartado 4 esta prevención “La ley limitará el uso de la informática para garantizar el honor y la intimidad personal y familiar de los ciudadanos y el pleno ejercicio de sus derechos”. Más que nunca se comprende la trascendencia

sistémica de estos impactos, y cómo se está jugando con fuego a nivel político y social.

- También a nivel individual, y de nuevo con repercusiones sociales sistémicas, la IA puede incrementar el impacto psicológico y cognitivo que ya están teniendo las tecnologías informáticas en general. El "entrenamiento" diario y masivo de jóvenes, mayores y niños en el uso de productos y servicios informáticos está generando importantes consecuencias a nivel cognitivo: pérdida de capacidad de atención y concentración⁴, rechazo progresivo de la interacción presencial o incluso meramente telefónica (oral), hiperemotivización del comportamiento (reacción por emoción), progresivo comportamiento no reflexivo, etc. Especialmente los más jóvenes acusan cambios de comportamiento respecto a las relaciones interpersonales y la comunicación, con especial mención de la reticencia a la expresión oral o la interacción presencial. Pero además con una creciente dedicación de tiempo a productos y servicios informáticos, en detrimento de experiencias interpersonales, con preocupantes tendencias al uso excesivo e incluso la adicción. Por otro lado, la IA generativa y la aplicación de la IA a la automatización, supone pasar de usar la IA para solucionar problemas difíciles a que la IA haga cosas por nosotros. Así que adicionalmente a la incidencia en la atención y la concentración habrá un impacto cognitivo más profundo pues "hacer" es el mecanismo más potente para aprender. Por ejemplo, si hasta ahora los menores leían poco y eran reticentes a leer o escribir en papel, o a la redacción de trabajos escritos, la aparición de ChatGPT ha reducido aún más estas actividades. ¿Cómo incidirá en las capacidades básicas de manejo y creación de documentos escritos que sustentan la actividad laboral y social en general? ¿Y en la creatividad? ¿Y en la actitud psicológica de una personalidad en construcción? Todo ello está alimentando progresivamente una cierta pérdida de la capacidad de reflexión. Un aspecto

4 ¿Por qué hemos perdido cuatro segundos de capacidad de atención en 15 años? *BBC News*. Disponible en: https://www.bbc.com/mundo/noticias/2016/02/160229_tecnologia_concentracion_distraccion_atencion_mz

fuertemente conectado con la autonomía personal y por lo tanto con la dinámica política y social básica para una sociedad abierta sostenible. Por si fuera poco, la IA, y especialmente su integración con avatares en entornos virtuales o videojuegos, puede amplificar enormemente este impacto cognitivo, incluso a nivel psicológico en general, al crear entes virtuales percibidos como una suerte de “seres” con los que pueden generarse lazos emocionales⁵.

- Finalmente, y enlazándolo con las reflexiones anteriores, la IA puede incrementar las capacidades de control de las personas, y su uso a nivel político o social. Aunque el Reglamento de IA recoge el scoring social entre las prácticas prohibidas, la IA permite múltiples posibilidades que inciden en el control ciudadano sin llegar a ser scoring social. Por otro lado, las prácticas de control social pueden ser también implementadas con otras tecnologías informáticas...

A estos y otros múltiples retos nos enfrenta el impacto de la IA a nivel político, ciudadano y social en general.

REFERENCIAS BIBLIOGRÁFICAS:

1. AFI: Analistas Financieros Internacionales (2024) Impacto de la IA en la economía española. Ocupaciones, sectores y CC.AA. Madrid. Disponible en: <https://www.afi.es/2024/05/20/inteligencia-artificial-economia-sectores-ocupaciones-comunidades-autonomas/> (Consulta: 10 de septiembre de 2024)
2. AMTEGA: Axencia para a Modernización Tecnolóxica de Galicia (2023) Guía práctica para la gestión de la inteligencia artificial en las administraciones públicas. Xunta de Galicia. Axencia para a Modernización Tecnolóxica de Galicia – Amtega. Santiago de Compostela. Disponible en:

5 Un japonés se casa con un holograma virtual. Diario ABC. 14 de noviembre de 2028. Disponible en: https://www.abc.es/internacional/abci-japones-casa-holograma-virtual-201811141406_video.html

- <https://amtega.xunta.gal/es/guia-practica-para-la-gestion-de-la-IA-en-las-AAPP> (Consulta: 10 de septiembre de 2024)
3. Comisión Europea (2024a) Public Sector Tech Watch: Mapping Innovation in the EU Public Services. A collective effort in exploring the applications of Artificial Intelligence and Blockchain in the Public Sector. Publications Office of the European Union, Luxembourg, 2024, ISBN 978-92-68-11627-2, doi: 10.2799/4393
 4. Manzoni, M., Medaglia, R., Tangi, L., Van Noordt, C., Vaccari, L. and Gattwinkel, D. (2022) AI Watch Road to the adoption of Artificial Intelligence by the Public Sector: A Handbook for Policymakers, Public Administrations and Relevant Stakeholders, EUR 31054 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-52131-0, doi:10.2760/693531, JRC129100.
 5. Peñarrubia, Juan Pablo (2024) ¿Quién cuida el jardín?: Cuidado humanizador de los cimientos individuales y sociales ante el impacto de las tecnologías digitales *SCIO. Revista de Filosofía*, n.º 26, Julio de 2024, 123-149, ISSN: 1887-9853 Disponible en: DOI: https://doi.org/10.46583/scio_2024.26.1155 (Consulta: 10 de septiembre de 2024)
 6. Randstad (2024) IA y mercado de trabajo en España. Randstad. Madrid. Disponible en: <https://www.randstadresearch.es/ia-mercado-trabajo-espana/> (Consulta: 10 de septiembre de 2024)
 7. Tangi L., van Noordt C., Combetto M., Gattwinkel D., Pignatelli F., (2022) AI Watch. European Landscape on the Use of Artificial Intelligence by the Public Sector. Joint Research Centre Report. EUR 31088 EN. Publications Office of the European Union, Luxembourg, ISBN 978-92-76-53058-9. doi:10.2760/39336, JRC129301.
 8. WEF: World Economic Forum (2023) Future of Jobs Report 2023. World Economic Forum). Ginebra. ISBN-13: 978-2-940631-96-4. Disponible en: <https://www.weforum.org/publications/the-future-of-jobs-report-2023/> (Consulta: 10 de septiembre de 2024). 

REGULACIÓN DE LA IA

🗨 José Manuel Muñoz Vela

INTRODUCCIÓN

La IA constituye un concepto complejo sobre el que no existe consenso ni científico, ni jurídico ni ético alrededor de su definición, si bien, los necesarios esfuerzos internacionales de armonización han adoptado de manera mayoritaria la última definición emanada de la OCDE.

La IA integra un conjunto de tecnologías y técnicas asociadas de incuestionable valor, impacto y potencial disruptivo y habilitador para cualquier sector y actividad, si bien, comporta diversos retos y riesgos que deben ser adecuadamente identificados y gestionados, y para los que los ordenamientos jurídicos vigentes no pueden ofrecer una adecuada respuesta, en la medida que fueron concebidos en un contexto tecnológico, político, jurídico, económico y social muy distinto al actual.

La evidencia y materialización de sus principales riesgos durante los últimos dos años, especialmente con la eclosión del despliegue y uso masivo de la IA generativa, ha acelerado e intensificado los debates internacionales sobre su seguridad y la necesidad de su regulación, ante la insuficiencia de la ética.

Las estrategias internacionales frente a la misma son muy diversas, al igual que sobre la necesidad de su regulación, encontrándonos países que consideran que no debe regularse, otros que debe hacerse, pero bajo un enfoque pro-innovación como Reino Unido o Israel y, otros que debe regularse con profundidad desde un enfoque global y de riesgos para garantizar su seguridad, su fiabilidad y el respeto de los derechos fundamentales, como la UE.

REGULACIÓN LEGAL, CONTRACTUAL Y CORPORATIVA

La IA no es una tecnología de reciente irrupción, en la medida que convivimos con la misma desde mediados del siglo pasado, si bien, es ahora cuando la mayor capacidad de computación, disponibilidad de datos y de otras

capacidades, ha permitido su evolución, eclosión y despliegue a un ritmo imparable, con la correlativa evidencia y materialización de sus riesgos asociados.

Durante sus aproximadamente 70 años de vida, la IA ha pasado por períodos de luces y de sombras, de grandes expectativas o de desmesuradas predicciones, con las correlativas desilusiones alrededor de la misma.

Su incesante desarrollo, despliegue y aplicación durante los últimos años planteó la necesidad de abordar sus retos y sus riesgos mediante distintos instrumentos, principalmente la ética y el Derecho, especialmente para garantizar su seguridad y la protección de los derechos fundamentales.

La ética se ha mostrado insuficiente hasta la fecha para garantizar todo ello, con frecuentes prácticas empresariales de *Ethic Washing*, de modo que, ante sus retos y riesgos, el carácter no imperativo de los principios y normas éticas esenciales objeto de mayor consenso internacional, así como la ausencia de una regulación legal de la misma hasta fechas recientes –que, entre otros aspectos, convierta dichos principios y normas éticas esenciales en imperativas–, hemos venido ordenando jurídicamente la misma a través de contratos entre todos los sujetos participantes en su ciclo de vida y cadena de valor, así como mediante códigos y normas corporativas vinculantes, creando así las bases de una incipiente rama jurídica transversal, como lo es, el Derecho de la IA.

El pasado 12 de julio de 2024, se publicó en el Diario Oficial de la UE el Reglamento IA de la UE, la primera norma en el mundo que pretende regular la IA de una manera global, horizontal y desde un enfoque de riesgos. El Reglamento entró en vigor el pasado 1 de agosto, si bien, su aplicación será diferida en el tiempo, como posteriormente se expone.

El Derecho de la IA que debe ordenar jurídicamente su desarrollo, despliegue, comercialización y uso debe estar integrado tanto por el denominado *hard law*, de carácter imperativo como el Reglamento IA de la UE, como por el denominado *soft law*, integrado por marcos de gobierno corporativos o sectoriales, códigos de conducta y estándares.

REGULACIÓN LEGAL EN LA UE

El *Libro blanco sobre la inteligencia artificial* elaborado por la Comisión Europea, de 19.02.2020, estableció la hoja regulatoria de la UE sobre la IA, basándose en un enfoque de riesgos y en una IA centrada en el ser humano, así como en la ética, la fiabilidad, la seguridad, la sostenibilidad y el respeto de los valores europeos y de los derechos fundamentales. Del mismo modo, clasificó los principales riesgos de la misma en tres categorías, esto es, riesgos para los derechos fundamentales, riesgos para la seguridad y riesgos para el funcionamiento eficaz del régimen de responsabilidad civil.

El Parlamento Europeo, efectuó dos recomendaciones a la Comisión sobre su regulación el 20 de octubre de 2020, de un lado, la *Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas*, que incorporó ya una propuesta reguladora del mismo a la Comisión, en particular, la *Propuesta de Reglamento del Parlamento europeo y del Consejo sobre los principios éticos para el desarrollo, el despliegue y el uso de la inteligencia artificial, la robótica y las tecnologías conexas*. En la misma, el Parlamento Europeo propuso a la Comisión la regulación de la IA mediante un instrumento jurídico de alcance general y de eficacia directa, que constituyó la primera propuesta reguladora a nivel europeo desde la ética y para la ética, que pretendía plasmar la visión reguladora y ética hasta entonces de la Comisión Europea sobre la IA que había sido recogida, entre otros documentos, en el precitado *Libro blanco sobre inteligencia artificial*, como en los trabajos previos del Grupo de Expertos de Alto Nivel sobre inteligencia artificial -AI HLEG- y en las *Directrices éticas para una inteligencia artificial fiable* de 2019. El instrumento propuesto tenía como principal objetivo garantizar un conjunto de principios y normas éticas esenciales, hasta entonces no vinculantes, y convertirlos en exigencias legales para todo sistema de IA, junto con otras adicionales y específicas para los sistemas considerados de alto riesgo, además de establecer un conjunto de obligaciones jurídicas adicionales para los mismos.

De otro, la *Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un régimen de*

responsabilidad civil en materia de inteligencia artificial, en la que incorporó igualmente una *Propuesta de Reglamento del Parlamento Europeo y del Consejo relativo a la responsabilidad civil por el funcionamiento de los sistemas de inteligencia artificial*. En la misma, el Parlamento propuso a la Comisión el establecimiento de un régimen legal de responsabilidad civil en materia de IA, mediante un instrumento jurídico con alcance general y de eficacia directa, que preveía un régimen de responsabilidad dual, esto es, de un lado, una responsabilidad objetiva para los sistemas clasificables conforme al mismo de alto riesgo y, de otro, una responsabilidad subjetiva para el resto, junto con otros aspectos como la responsabilidad solidaria en caso de pluralidad de agentes que pudieran intervenir en el ciclo de vida de un sistema inteligente.

Sin embargo, las propuestas dirigidas por el Parlamento Europeo a la Comisión en estas cuestiones no derivaron en las correlativas iniciativas legislativas por parte de ésta.

La Comisión publicó el 21 de abril de 2021, la primera versión del Reglamento IA de la UE que, tras tres intensos años de tramitación, con duras negociaciones y sucesivas enmiendas, fue finalmente aprobado y publicado el pasado 12 de julio de 2024¹.

El Reglamento IA de la UE pretende constituir una regulación global y horizontal de la IA (no vertical y sectorial) desde un enfoque de riesgos, el cual, desde el inicio de su tramitación, se constituyó en la norma de referencia internacional ante el denominado "efecto Bruselas".

Se trata de una norma de primera generación que no es perfecta ni completa, pero que pretende ofrecer una regulación proactiva y *ex ante* de la IA ante sus riesgos, con una pretendida eficacia universal elevada (extraterritorial) desde su concepción.

No obstante, el Reglamento IA de la UE, no tiene realmente el alcance global con el que se presenta, en la medida que el mismo, de un lado, se focaliza

1 Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) nº 300/2008, (UE) nº 167/2013, (UE) nº 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial). DOUE Nº. 1689. 12.07.2024. Pp. 1-144.

en prohibir determinados sistemas por considerarlos de riesgo inadmisibles² y en regular con intensidad y minuciosidad técnica los requisitos de los sistemas considerados conforme al mismo de alto riesgo³, así como las obligacio-

-
- 2 Sistemas o aplicaciones de IA que sean destinados o que provoquen manipular subliminalmente el comportamiento humano de forma que se cause un daño a sí mismo o a terceros; sistemas diseñados o que se aprovechen de la vulnerabilidad de una persona o grupo por razones de edad, discapacidad física o psíquica, situación social o económica específica, para manipular su comportamiento de modo sustancial provocando un perjuicio; sistemas que permiten la evaluación o clasificación de personas o grupos (puntuación o *scoring* social) con fines públicos o privados; sistemas para la evaluación o predicción de la probabilidad de comisión de delitos por parte de una persona basados exclusivamente en la elaboración de su perfil o de los rasgos y características de su personalidad (salvo para apoyar valoración humana de la implicación de una persona en una actividad delictiva); sistemas de categorización biométrica que clasifiquen individualmente a personas físicas basándose en sus datos biométricos para deducir o inferir su raza, opiniones políticas, afiliación sindical, creencias religiosas o filosóficas, vida sexual u orientación sexual; sistemas de identificación biométrica remota en tiempo real y en espacios de acceso público con fines de aplicación de la ley, con excepciones; sistemas de IA que creen o amplíen bases de datos de reconocimiento facial mediante el *scraping* no selectivo de imágenes faciales de Internet o de grabaciones de CCTV; sistemas de IA para inferir emociones de una persona física en los ámbitos del lugar de trabajo y de las instituciones educativas (excepto por razones médicas o de seguridad); otros. Significar que los sistemas, expresa y específicamente incluidos y prohibidos en el Reglamento no son los únicos, sino que, del mismo modo, estará igualmente prohibido cualquier otro sistema que pueda infringir otros marcos jurídicos aplicables, en la medida que, tal y como establece el apartado 8 de su artículo 5 “el presente artículo no afectará a las prohibiciones aplicables cuando una práctica de IA infrinja otras disposiciones de Derecho de la Unión”.
- 3 Sistemas de IA que estén destinados a ser utilizados como componente de seguridad de un producto sujeto a la legislación armonizada de seguridad de los productos de la UE relacionada en el Anexo I o sistemas que sean el producto en sí mismo sujeto a ésta, siempre y cuando dicho producto del que el sistema de IA sea componente de seguridad o el propio sistema de IA como producto sujeto a este marco legislativo deba someterse a una evaluación de la conformidad de terceros para su introducción en el mercado o puesta en servicio –Juguetes, máquinas, aviación civil, embarcaciones de recreo, motos acuáticas, aparatos y sistemas de protección para uso en atmósferas potencialmente explosivas, equipos radioeléctricos, equipos a presión, instalaciones de transporte por cable, equipos de protección individual, aparatos para la quema de combustibles gaseosos, productos sanitarios (incluidos para diagnóstico in vitro), vehículos de dos, tres o cuatro

nes (pre y poscomercialización) de los distintos sujetos relacionados con los mismos, esto es, los operadores -entre los que se incluyen los proveedores, responsables del despliegue ("usuarios" en su redacción primigenia y referido a los profesionales y entidades públicas o privadas que utilicen sistemas inteligentes), importadores, distribuidores, fabricantes de productos que integren sistemas inteligentes y representantes autorizados-.

Del mismo modo, regula los modelos y sistemas de IA de uso general con/sin riesgo sistémico, así como determinadas obligaciones específicas de información y transparencia para concretos sistemas de IA al margen de su clasificación en las categorías anteriores y de riesgo limitado (Por ejemplo, sistemas inteligentes que interaccionen con personas).

ruedas, vehículos agrícolas o forestales, equipos marinos, dispositivos médicos o ascensores-. Así como los sistemas expresamente clasificados de alto riesgo para la salud y la seguridad o los derechos fundamentales por el Reglamento (Revisables, ampliables y con excepciones), que incluyen: a) Sistemas de identificación biométrica remota (con excepciones); b) sistemas de IA destinados a ser utilizados para la categorización biométrica en función de atributos o características sensibles o protegidos basada en la inferencia de dichos atributos o características; c) sistemas de IA destinados a ser utilizados para el reconocimiento de emociones; d) sistemas destinados a ser utilizados como componentes de seguridad en la gestión y funcionamiento de infraestructuras digitales críticas, de tráfico rodado o del suministro de agua, gas, calefacción o electricidad; e) sistemas a ser utilizados en educación y formación profesional con determinadas finalidades; f) sistemas a ser utilizados en materia de empleo, gestión de los trabajadores y acceso al autoempleo con determinadas finalidades; g) sistemas a ser utilizados para determinadas finalidades en ámbito del acceso y utilización de servicios privados y públicos esenciales o prestaciones públicas esenciales (incluye asistencia sanitaria, calificación crediticia o evaluación de riesgos y tarificación en el sector de los seguros); h) sistemas a ser utilizados con determinadas finalidades por las autoridades garantes del cumplimiento del Derecho o entidades, órganos u organismos en apoyo de éstas; i) sistemas a ser utilizados para la gestión de migración, asilo y control de fronteras con determinadas finalidades (siempre que su uso esté permitido por el Derecho de la UE o nacional aplicable); j) sistemas a ser utilizados en el ámbito de la Administración de justicia y en procesos democráticos con concretas finalidades. No obstante, el Reglamento prevé excepciones a la clasificación de estos sistemas como de alto riesgo cuando no planteen un riesgo importante de causar un perjuicio a la salud, la seguridad o los derechos fundamentales de las personas físicas.

Salvo las excepciones indicadas, el Reglamento IA de la UE no regula el resto de sistemas (riesgo limitado o mínimo), para los que promueve la autorregulación dispositiva (códigos de conducta).

El Reglamento IA de la UE no se aplica a sistemas con fines militares, de defensa o seguridad nacional, ni a modelos y sistemas con finalidades científicas y de investigación o en fase de prueba de manera previa a su introducción en el mercado, ni a sistemas distribuidos bajo licencias de software libre o código abierto salvo excepciones, entre otros.

El objeto y alcance específicos de esta guía impide abordar con profundidad los distintos aspectos que comporta el Reglamento, si bien, en otros aspectos a destacar con los que he sido muy crítico desde la publicación de la primera versión el 21 de abril de 2021, significar que el Reglamento IA de la UE no regula y exige los principios y normas éticas más esenciales y esperables para cualquier sistema inteligente que se introduzca en el mercado o use, como seguridad, intervención y vigilancia humanas, al servicio de las personas, respeto de la dignidad humana y la autonomía personal, solidez, privacidad, gobernanza y calidad de los datos, transparencia (trazabilidad y explicabilidad adecuadas), diversidad, no discriminación y equidad, bienestar social y medioambiental (sostenibilidad). Incongruente para un instrumento jurídico tan ambicioso en su pretensión y elaborado desde un enfoque de riesgos.

Durante su tramitación, se incorporó a los debates y negociaciones la inclusión de un nuevo artículo 4bis que contemplaba estos principios como exigencia jurídica para cualquier sistema inteligente, si bien, debido a la presión de las partes interesadas, fue finalmente eliminado del texto finalmente aprobado.

El Reglamento IA de la UE también establece la necesidad de medidas de apoyo para su cumplimiento para *pymes* y *startups*, regula y fomenta los espacios controlados de pruebas (*sandbox*) y promueve el derecho de los ciudadanos a presentar quejas y recibir explicaciones, la alfabetización en IA, así como la normalización y la estandarización como instrumentos para facilitar el cumplimiento del mismo, la conformidad y la armonización.

El Reglamento establece un marco de gobernanza, integrado por las autoridades nacionales de supervisión, el Comité Europeo de IA, la Oficina Europea de IA, el foro consultivo y la comisión científica de expertos.

Por último, regula un marco sancionador en el que las sanciones más graves podrán llegar hasta los 35.000.000€ o el 7 % del volumen de negocios total anual de la empresa infractora a escala mundial correspondiente al ejercicio anterior, si esta cifra es superior.

El Reglamento precisará distintas acciones de acompañamiento y de desarrollo de sus disposiciones mediante actos delegados de la Comisión y de ejecución, directrices, guías y modelos de la Oficina Europea de IA y de las agencias de supervisión de IA nacionales.

Su aplicación y exigibilidad será escalonada. Tras su publicación en el Diario Oficial de la UE el 12 de julio de 2024, entró en vigor el 1 de agosto de 2024, si bien, sus disposiciones no serán aplicables hasta el 2 de agosto de 2026, con excepciones⁴.

Por último, el Reglamento IA de la UE no contempla en sus disposiciones la regulación de la responsabilidad civil por daños derivados del funcionamiento y uso de la IA, que era uno de los riesgos identificados en el precitado *Libro blanco sobre IA*, si bien, incorpora distintas disposiciones reguladoras de especial relevancia para su determinación y exigencia.

La Comisión optó por regular esta materia mediante la revisión y adecuación de los regímenes vigentes reguladores de la misma a la IA, mediante un instrumento jurídico diferenciado al Reglamento IA de la UE, si bien, sorpresivamente, lo hizo mediante la publicación de la *Propuesta de Directiva del Parlamento Europeo y del Consejo, relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial (Directiva sobre la responsabilidad en materia de IA)*, de 28 de septiembre de 2022, la cual fue acompañada por la coetánea *Propuesta de Directiva del Parlamento Europeo y del Consejo, sobre responsabilidad por los daños causados por*

⁴ Los capítulos I y II serán aplicables a partir del 2 de febrero de 2025; el capítulo III, sección 4, el capítulo V, el capítulo VII y el capítulo XII y el artículo 78 serán aplicables a partir del 2 de agosto de 2025, a excepción del artículo 101; el artículo 6, apartado 1, y las obligaciones reguladas en el Reglamento serán aplicables a partir del 2 de agosto de 2027

productos defectuosos, también de 28 de septiembre de 2022, que pretende derogar la Directiva 85/374/CEE del Consejo, de 25 de julio de 1985, sobre responsabilidad por los daños causados por productos defectuosos.

La tramitación de la primera de las Directivas propuestas no ha avanzado durante estos dos años, produciéndose la aprobación del Reglamento IA de la UE por el camino que motivó la revisión de su texto para adecuarlo al mismo por parte de la Comisión, mientras que el texto de la segunda fue aprobada por el Parlamento Europeo en marzo de 2024, hallándose pendiente de aprobación formal por el Consejo en la fecha de cierre de este monográfico, para su posterior entrada en vigor a los 20 días de su publicación en el Diario Oficial de la UE.

REGULACIÓN LEGAL INTERNACIONAL: PROPUESTAS

La eclosión del despliegue y uso de la IA generativa, junto con la materialización y evidencia de los riesgos asociados a la IA en general que se venían anunciando, como se ha referido anteriormente, intensificaron los debates alrededor de su seguridad y sobre la necesidad de su regulación, cuando menos para garantizar la salud, la seguridad y los derechos fundamentales.

Durante estos últimos años las iniciativas reguladoras internacionales han proliferado, si bien, bajo distintas estrategias y enfoques, significando los países más focalizados en la innovación y la autorregulación como Reino Unido o Israel, frente a otros que han adoptado un enfoque similar al de la UE. Por su parte, como se expondrá a continuación, EE.UU. ha adoptado una postura prudente ante el potencial mundial de su industria nacional pero firme, apostando por el diálogo y compromiso gubernamental y agencias con la propia industria, y su acompañamiento por acciones ejecutivas y legislativas.

España dispone de distintas normas reguladoras de determinados aspectos relacionados con la IA, tanto a nivel estatal como autonómico, entre otros: el *Real Decreto-ley 9/2021, de 11 de mayo, por el que se modifica el texto refundido de la Ley del Estatuto de los Trabajadores, aprobado por el Real Decreto Legislativo 2/2015, de 23 de octubre, para garantizar los derechos laborales de las personas dedicadas al reparto en el ámbito de plataformas digitales* (Ley Riders); la *Ley 1/2022, de 13 de abril, de Transparencia y Buen Gobierno de la Comunidad Valenciana*; la *Ley 15/2022, de 12 de julio, integral*

para la igualdad de trato y la no discriminación, o; el Decreto-ley 2/2023, de 8 de marzo, de medidas urgentes de impulso a la IA en Extremadura. Del mismo modo, España se ha posicionado a la vanguardia en materia de regulación y gobierno de la IA, siendo el primer país en la UE en crear su Agencia de Supervisión de la IA, mediante la *Ley 28/2022, de 21 de diciembre, de fomento del ecosistema de las empresas emergentes* (Disposición adicional séptima), a la que prosiguió la norma que regula su estatuto y funciones, esto es, *el Real Decreto 729/2023, de 22 de agosto, por el que se aprueba el Estatuto de la Agencia Española de Supervisión de Inteligencia Artificial*, siendo igualmente pionera en la creación de un *sandbox* regulatorio de IA mediante *Real Decreto 817/2023, de 8 de noviembre, que establece un entorno controlado de pruebas para el ensayo del cumplimiento de la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial*. Asimismo, mediante Orden TDF/619/2024, de 18 de junio de 2024, ha creado y regulado el Consejo Asesor Internacional en Inteligencia Artificial, y está tramitando una Proposición de Ley Orgánica, de 13 de octubre de 2023, de regulación de las simulaciones de imágenes y voces de personas generadas por medio de la inteligencia artificial.

Por lo que se refiere a EE.UU. al margen de los diálogos y compromisos alcanzados con los gigantes de IA estadounidenses durante 2023, significar la Orden Ejecutiva 14110 sobre una IA segura y fiable precitada, publicada bajo el título *Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, la cual establece la hoja de ruta ejecutiva y de acciones legislativas de acompañamiento en materia de IA, con el objetivo de garantizar la gobernanza del desarrollo y uso de la IA de manera segura y responsable, impulsando un enfoque coordinado a nivel de todo el gobierno federal para hacerlo y que se ha venido cumpliendo de manera exhaustiva desde su publicación.

Del mismo modo, junto las iniciativas federales, se han aprobado y se están también tramitando distintas iniciativas estatales y locales en todo el país, como California, Connecticut, Florida, Luisiana, Minnesota, Montana, Texas, Virginia, Illinois, Alabama, Washington, Nueva York, Vermont, Tennessee o Colorado. A nivel local, Boston, Miami, Nueva York, San José (California) o Seattle también han creado marcos reguladores y directrices sobre diversos

aspectos de la IA, como la IA generativa, los sistemas automatizados de decisión laboral y las evaluaciones de impacto.

Como ejemplo local, Boston publicó sus orientaciones sobre IA generativa, que incluyen principios sobre precisión y transparencia. Washington DC tramitó en 2021 una propuesta contra la no discriminación bajo el título *Stop Discrimination by Algorithms Act*. Seattle adoptó una política de IA generativa en mayo de 2023 que ha sido tomada como referencia para la propuesta regulatoria de la IA generativa en el estado de Washington.

Canadá, Chile, Brasil, Colombia, Argentina, México, Perú, Reino Unido, Turquía, Kenia, Nigeria, Egipto, Rusia, China, Singapur, Corea del Sur o Vietnam, están tramitando en la fecha de cierre de esta guía sus iniciativas reguladoras bajo diversos enfoques, objeto, alcance y técnicas legislativas para su consecución.

BIBLIOGRAFÍA

1. BARRIO ANDRÉS, M. (2023). "Novedades en la tramitación del próximo Reglamento europeo de inteligencia artificial". *ARI Real Instituto Elcano*. N. 67/2023. Madrid.
2. MUÑOZ VELA, J.M. (2021). *Cuestiones éticas de la Inteligencia Artificial y repercusiones jurídicas. De lo dispositivo a lo imperativo*. Ed. Aranzadi Thomson Reuters. Navarra.
3. MUÑOZ VELA, J.M. (2022). "Inteligencia artificial y responsabilidad penal". *Derecho Digital e Innovación*. Núm. 11. Enero-Marzo 2022. ISSN 2659-871X, Ed. Wolters Kluwer.
4. MUÑOZ VELA, J.M. (2022). *Retos, riesgos, responsabilidad y regulación de la inteligencia artificial. Un enfoque de seguridad física, lógica, moral y jurídica*. Editorial Aranzadi Thomson Reuters. Navarra.
5. MUÑOZ VELA, J.M. (2022). "Inteligencia artificial y Cuestiones de propiedad intelectual e industrial". *Revista Aranzadi de derecho y nuevas tecnologías*. ISSN 1696-0351, Nº. 59. Editorial Aranzadi (Thomson Reuters).
6. MUÑOZ VELA, J.M. (2022). "Avanzando en la regulación de la IA. Hacia un equilibrio entre ética, protección de los derechos fundamentales, seguridad, confianza, innovación, desarrollo y competitividad". *Derecho*

- Digital e Innovación*. N. 14. Octubre-Diciembre. 2022. ISSN 2659-871X, Ed. Wolters Kluwer.
7. MUÑOZ VELA, J.M. (2023) "Inteligencia artificial y derecho de autor. Creaciones artificiales y su protección jurídica", en Martínez Velencoso, L.M. y Plaza Penadés, J. (Dir.) (2023), *Retos normativos del mercado único digital europeo*. Tirant Lo Blanch.
 8. MUÑOZ VELA, J.M. (2023). "IA y responsabilidad civil. Comentarios a las propuestas europeas en materia de derechos de daños por productos defectuosos y adaptación de las normas de responsabilidad civil extracontractual". *Revista Aranzadi de derecho y nuevas tecnologías*. N. 61. Editorial Aranzadi Thomson Reuters.
 9. MUÑOZ VELA, J.M. (2024). "Inteligencia artificial generativa. Desafíos para la propiedad intelectual". *Revista de Derecho UNED*. N. 33. Primer premio García Goyena.
 10. MUÑOZ VELA, J.M. (2024). *La regulación de la inteligencia artificial. Reto y oportunidad desde una perspectiva global e internacional*. Ed. Aranzadi Thomson Reuters. Navarra.
 11. PEGUERA POCH, M. (2023). "La propuesta de Reglamento de IA: una intervención legislativa insoslayable en contexto de incertidumbre", en PEGUERA POCH, M. (Coord.) *Perspectivas regulatorias de la inteligencia artificial en la Unión Europea*. Editorial Reus. Madrid. 

GOBERNANZA DE LA IA

💬 Juan Pablo Peñarrubia Carrión

La gobernanza de la IA tiene, al menos, dos niveles de interpretación: A nivel empresarial, es decir, en la dirección y acción de una empresa u organización en general para la consecución de sus objetivos; y a nivel social, es decir, en la acción de los poderes públicos (gobiernos, legisladores, justicia...), las personas y los actores sociales (empresas, asociaciones, organizaciones de todo tipo etc.). En este segundo caso hay además al menos dos escalas relevantes: el nivel de los estados y el nivel internacional o global, y en su materialización tiene especial relevancia la visión adoptada para la regulación y el impacto ético. Conviene esta reflexión para comprender a que nos referimos cuando hablamos de gobernanza de la IA.

A nivel social no existe una gobernanza de la IA a nivel global. Pero, a pesar de la ausencia regulatoria, la incertidumbre y de algunas llamadas a una moratoria en el uso de los sistemas de IA en 2023, la sociedad no va a esperar a la era de la regulación digital dura, ni a deseables consensos regulatorios, ni mucho menos a consensos éticos. (Domingo y Peñarrubia, 2024b). Las diferencias regulatorias, éticas y culturales generan diferentes aproximaciones de gobernanza de la IA en los estados. En el epígrafe sobre ética y regulación se perfilan aspectos de las visiones de la IA más relevantes a escala global: Unión Europea, EEUU y China.

Este documento no permite un análisis detallado de todas las vertientes de la gobernanza de la IA, por lo que siguiendo el objetivo orientador de la publicación nos centraremos en la visión europea de la gobernanza de la IA como sociedad y la gobernanza de la IA a nivel de las empresas y organizaciones en general.

VISIÓN EUROPEA DE LA GOBERNANZA DE LA IA

La gobernanza de la IA en la Unión Europea está basada en la visión regulatoria de la IA con dos componentes: la normativa legal, con especial mención del Reglamento de IA, y la normativa de estandarización en IA.

La materialización de la gobernanza de la IA se realiza en un ciclo con tres elementos esenciales: las disposiciones de la regulación normativa de IA (tanto normas legales como normas de estandarización), la acción de las autoridades de control del cumplimiento normativo, y la acción de los actores sociales: ciudadanía y organizaciones en general en su diferentes roles de IA (ISO, 2022a), incluyendo empresas y entidades del sector público. Dicha acción, a su vez retroalimenta la mejora y actualización de la regulación normativa, tanto legal como de estandarización.

Comprender este modelo de gobernanza es esencial para llevarlo a la práctica de un modo viable y valioso para las empresas y organizaciones en general, especialmente cuando se trate de productoras de IA o proveedoras de IA. Y en esa puesta en práctica el elemento clave para las organizaciones serán las normativas de estandarización de IA, entre otros motivos porque:

- El cumplimiento de normativas de estandarización se considera presunción de conformidad de requisitos establecidos en el Reglamento de IA. Es decir, la normativa de estandarización detalla a las organizaciones y profesionales como implementar de modo práctico las obligaciones genéricas establecidas en el Reglamento de IA, tanto a nivel tecnológico, organizativo y documental.
- Recogen buenas prácticas de la industria y los profesionales, lo que re-
donda en la confiabilidad de los correspondientes sistemas de IA y en la mejor gobernanza de la IA a nivel de las organizaciones.
- Orientan la puesta en práctica de medidas en relación a sistemas de IA incluso cuando no son requisito legal, sino por usos normalizados, lo que aumentará la productividad, generará sistemas de IA de mayor calidad y facilitará la integración de diferentes sistemas y la movilidad de los profesionales conocedores de dichas prácticas normalizadas.

Para materializar este modelo de gobernanza la Comisión Europea formula "peticiones de normalización" (*standardisation request*) en los aspectos que considera necesario para complementar y llevar a la práctica las previsiones legales, incluyendo obligaciones, información, documentación, controles, etc. (Como se recoge en el propio Reglamento de IA). Así, en paralelo

a la elaboración del Reglamento de IA, la Comisión Europea realizó en mayo de 2023 una petición de normalización en IA (Comisión Europea, 2023), que recoge la primera de remesa de normativas de estandarización que se consideran prioritarias y esenciales para complementar lo dispuesto en el Reglamento de IA (Figura 1).

List of new European Standards and European standardisation deliverables to be drafted

Reference information	
1.	European standard(s) and/or European standardisation deliverable(s) on risk management systems for AI systems
2.	European standard(s) and/or European standardisation deliverable(s) on governance and quality of datasets used to build AI systems
3.	European standard(s) and/or European standardisation deliverable(s) on record keeping through logging capabilities by AI systems
4.	European standard(s) and/or European standardisation deliverable(s) on transparency and information provisions for users of AI systems
5.	European standard(s) and/or European standardisation deliverable(s) on human oversight of AI systems
6.	European standard(s) and/or European standardisation deliverable(s) on accuracy specifications for AI systems
7.	European standard(s) and/or European standardisation deliverable(s) on robustness specifications for AI systems
8.	European standard(s) and/or European standardisation deliverable(s) on cybersecurity specifications for AI systems
9.	European standard(s) and/or European standardisation deliverable(s) on quality management systems for providers of AI systems, including post-market monitoring processes
10.	European standard(s) and/or European standardisation deliverable(s) on conformity assessment for AI systems

Figura 1. Lista de normas y documentos de estandarización contenidos en la petición de normalización sobre IA de la Comisión Europea de mayo de 2023. Fuente (Comisión Europea, 2023)

Este conjunto de normas de estandarización es materializado por un Comité Técnico Conjunto CEN-CENELEC (European Committee for Standardisation - European Committee for Electrotechnical Standardisation), en concreto

el comité CEN/CLC/JTC 21 – Artificial Intelligence, en el que España está representada por la Asociación Española de Normalización (UNE). El cometido de este comité es: elaborar productos de normalización en el ámbito de la IA y orientar a otros comités técnicos en relación a la IA; considerar la adopción de normas internacionales pertinentes y normas de otras organizaciones pertinentes, como ISO/IEC JTC 1 y sus subcomités, como el SC 42 Inteligencia Artificial; y producir productos de normalización para responder a las necesidades del mercado europeo y de la sociedad y para respaldar principalmente la legislación, las políticas, los principios y los valores de la Unión Europea¹. El calendario asociado a la materialización de esta petición de estandarización está alineado con el calendario del Reglamento de IA, publicado en junio de 2024, pero que establece un calendario progresivo de entrada en vigor de las diferentes previsiones hasta 2026. El conjunto de estándares resultado de la petición de estandarización, y extractados en la Figura 1 estarán disponibles a lo largo de 2025 y 2026. A fecha de elaboración de esta publicación el comité CEN/CLC/JTC 21 ha publicado 7 documentos de estandarización y tiene 31 en curso.

En cuanto a las autoridades para la gobernanza de la IA en la Unión Europea, el Reglamento de IA especifica las autoridades de control del cumplimiento legal en materia de IA y algunos órganos complementarios para la gobernanza de la IA en la Unión Europea (Comisión Europea, 2024b):

- La Oficina Europea de Inteligencia Artificial, que tiene como misión desarrollar conocimientos especializados y capacidades de la Unión Europea en el ámbito de la IA y contribuir a la aplicación de la legislación europea sobre IA.
- Consejo Europeo de Inteligencia Artificial, compuesto por representantes de los Estados miembros, debe apoyar a la Comisión para promover herramientas de alfabetización en materia de IA, la sensibilización pública

1 Espacio web de CEN/CLC/JTC 21 – Artificial Intelligence: https://standards.cencenelec.eu/dyn/www/f?p=205:22:0:::FSP_ORG_ID,FSP_LANG_ID:2916257,25&cs=1827B89DA69577B-F3631EE2B6070F207D

y la comprensión de los beneficios, los riesgos, las salvaguardias, los derechos y las obligaciones en relación con el uso de sistemas de IA

- Grupo de expertos científicos independientes, destinado a apoyar las actividades de garantía del cumplimiento en su vertiente más técnica o ante futuras innovaciones.
- Foro consultivo para facilitar las aportaciones de las partes interesadas por la aplicación de la normativa europea de IA. Incluyendo representantes de las organizaciones europeas de estandarización.
- Autoridades nacionales competentes. Cada Estado miembro dispondrá de al menos una autoridad notificante y al menos una autoridad de vigilancia del mercado como autoridades nacionales competentes en IA

Sobre la gobernanza de la IA en las empresas y organizaciones en general, es esencial entender que las organizaciones necesitan comprender y concretar, a nivel técnico, material y organizativo, como realizar la gobernanza de la IA. Aunque ello es especialmente relevante para los productores de IA y los proveedores de IA, debe ser igualmente considerado por los usuarios de IA o cualquier otro rol de IA. El uso de la IA ha de ser necesariamente gestionado e integrado en el conjunto de la gestión y la gobernanza de las organizaciones, no solo por conveniencia organizativa sino también por las potenciales responsabilidades y diligencia debida con relación a las peculiaridades de la IA. Y tanto a nivel directivo, como de la empresa en su conjunto, tanto con relación a sus empleados y esfera interna, como a sus clientes, proveedores, consumidores y esfera externa en general. A continuación, se recogen algunas normas clave para la materialización de la gobernanza de la IA en las organizaciones, tanto por su trascendencia intrínseca, como por el hecho de estar en curso la publicación de las correspondientes versiones en español desde UNE:

- ISO/IEC 42001 Tecnologías de la información — Inteligencia artificial — Sistema de gestión (*ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system*) (ISO, 2023b): Especifica los requisitos y proporciona orientación para establecer, implementar, mantener y mejorar continuamente un sistema de gestión de la IA en el contexto

de una organización. Facilita a la organización desarrollar, proveer o usar sistemas de IA de forma responsable en la consecución de sus objetivos y a cumplir los requisitos aplicables, incluyendo la relación con las partes interesadas. Aplicable a cualquier organización, independientemente de su tamaño, tipo y naturaleza, que provea o use productos o servicios que utilicen sistemas de IA. La importancia de esta norma es central para la gobernanza de la IA en las organizaciones dado que aplica a todos los roles de IA y que al tratarse de un estándar de sistema de gestión, se integra de modo coherente con otros estándares de sistemas de gestión ampliamente conocidos y utilizados por las organizaciones (por ejemplo ISO 9001 Sistema de gestión de calidad, ISO 27001 Sistema de gestión de la seguridad de la información, etc.) El estándar a nivel global fue publicado en 2023. UNE está realizando el proyecto de versión en español, para lo cual encargó en 2024 la traducción de la norma original al Consejo General de Colegios de Ingeniería Informática de España (CCII), estando el proyecto actualmente en los siguientes estadios de tramitación formal, con estimación de publicación de la norma a finales de 2024 o principio de 2025.

- ISO/IEC 23894 Tecnologías de la información — Inteligencia artificial — Guía sobre gestión del riesgo (*ISO/IEC 23894:2023 Information technology — Artificial intelligence — Guidance on risk management*) (ISO, 2023a): Proporciona orientación sobre a las organizaciones que desarrollan, producen, despliegan o usan productos, sistemas y servicios que utilizan IA pueden gestionar el riesgo específicamente relacionado con la IA. También debe servir a las organizaciones para integrar la gestión del riesgo en sus actividades y funciones relacionadas con la IA. Se debe utilizar conjuntamente con la Norma UNE-ISO 31000:2018 Gestión del Riesgo. El estándar a nivel global fue publicado en 2023. UNE está realizando un proyecto de versión en español análogo al anterior, también con estimación de publicación a finales de 2024 o principio de 2025.
- ISO/IEC 38507 Tecnologías de la información — Gobernanza de Tecnologías de la Información — Implicaciones para la gobernanza del uso de la inteligencia artificial por las organizaciones (*ISO/IEC 38507:2022 Information technology — Governance of IT — Governance implications of the use*

of artificial intelligence by organizations) (ISO, 2022b): Proporciona orientación a los miembros de los órganos de gobierno de una organización para permitir y gobernar el uso de la IA, con el fin de garantizar su utilización eficaz, eficiente y aceptable dentro de la organización. También proporciona orientación a directivos, profesionales de cualquier ámbito, asociaciones empresariales, colegios profesionales, autoridades públicas, proveedores de servicios, auditores, etc. Aplicable a la gobernanza de los usos actuales y futuros de la IA, así como a las implicaciones de dicho uso para la propia organización. Aplicable a cualquier organización, y de cualquier tamaño, incluidas empresas públicas y privadas, entidades gubernamentales y organizaciones sin ánimo de lucro.

ÚLTIMAS NOVEDADES SOBRE GOBERNANZA DE LA IA A NIVEL GLOBAL :

A nivel de gobernanza de la IA a nivel global se están impulsando iniciativas para explorar formulas viables y efectivas. Cabe destacar por su reciente celebración el Foro Global UNESCO sobre la ética de la IA celebrado en Kranj (Eslovenia) con el título "Cambiando el panorama de la gobernanza de la IA" , en febrero de 2024, entre cuyas principales conclusiones está la importancia de la colaboración internacional. Y la Conferencia Mundial sobre IA (World Artificial Intelligence Conference – WAIC), celebrada en Shanghai en julio de 2024, y en cuyo marco se presentó la Declaración de Shanghai sobre Gobernanza Global de la IA. La declaración propone cinco líneas de acción para el impulso de la IA abierta y la cooperación internacional: promover el desarrollo de la IA; mantener la seguridad de la IA; desarrollar el sistema de gobernanza de la IA, tanto a nivel global como de los estados; incrementar la información y participación de la ciudadanía; mejorar la calidad de vida y el bienestar social.

REFERENCIAS BIBLIOGRÁFICAS:

1. Comisión Europea (2023) C(2023)3215 – Standardisation request M/593 commission implementing decision of 22.5.2023 on a standardisation request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in support of Union policy on artificial intelligence. Disponible en <https://ec.europa>.

- eu/growth/tools-databases/enorm/mandate/593_en (Consulta: 3 de septiembre de 2024)
2. Comisión Europea (2024b) Reglamento (UE) 2024/1689 Del Parlamento y del Consejo, de 13 de junio de 2024 por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.o 300/2008, (UE) n.o 167/2013, (UE) n.o 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial) Diario Oficial, L 1689, ELI. Disponible en: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj> (Consulta: 3 de septiembre de 2024)
 3. Domingo, Agustín; Peñarrubia, Juan Pablo (2024b) Faros en la niebla: Herramientas de ética aplicada para sistemas de IA. Conferencia en XXV World Congress of Philosophy: "Philosophy across Boundaries". 1-8 Agosto 2024. Roma. Sitio web del congreso: <https://wcprome2024.com>
 4. ISO (2022a) ISO/IEC 22989:2022 Information technology — Artificial intelligence — Artificial intelligence concepts and terminology. International Organization for Standardization – ISO. Ginebra. Disponible en: <https://www.iso.org> (Consulta: 8 de septiembre de 2024)
 5. ISO (2022b) ISO/IEC 38507:2022 Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations. International Organization for Standardization – ISO. Ginebra. Disponible en: <https://www.iso.org> (Consulta: 8 de septiembre de 2024)
 6. ISO (2023a) ISO/IEC 23894:2023 Information technology — Artificial intelligence — Guidance on risk management. International Organization for Standardization – ISO. Ginebra. Disponible en: <https://www.iso.org> (Consulta: 8 de septiembre de 2024)
 7. ISO (2023b) ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system. International Organization for Standardization – ISO. Ginebra. Disponible en: <https://www.iso.org> (Consulta: 8 de septiembre de 2024). 

IMPACTO DE LA IA EN LA SOCIEDAD DEL FUTURO

💬 Mar Monsoriu Flor

Esta *“Guía sobre Inteligencia Artificial”* (IA), en la que hemos colaborado profesionales de diferentes disciplinas, no estaría completa si no le dedicásemos un breve capítulo al futuro. En él se exponen aspectos positivos que derivarán del uso intensivo de la IA y los grandes desafíos a los que nos enfrentaremos como sociedad global.

Omar Hatamleh, jefe de Inteligencia Artificial del Centro Goddard de Vuelos Espaciales de la NASA en Maryland (EE.UU.), asegura que “la IA nos ayudará a imprimir órganos humanos con la misma genética de una persona para evitar rechazos. Además, tendremos gemelos digitales que permitirán que la medicina sea completamente individualizada. Se harán millones de simulaciones en un gemelo digital para determinar qué tratamiento es el más adecuado para cada caso. Incluso se podrán saber qué enfermedades sufrirá alguien. Lo que se traduce en que la vida se va a extender 100 o 120 años de media, con las implicaciones que tendrá esto en el mercado laboral y de jubilación”.

Otros expertos coinciden con Hatamleh tanto en el enorme desarrollo de la IA en la medicina, como en el enorme impacto que va a tener la IA en el empleo. La IA se está utilizando en el ámbito de Ciencias de la Salud (Medicina, Veterinaria, Biotecnología, Odontología, Farmacia, etcétera) porque se trata de un sector que afecta directamente a la vida de las personas y que además mueve mucho dinero. En los países desarrollados cuenta con una fuerte financiación y a él se destina un buen porcentaje de los presupuestos nacionales. Hay que añadir que tanto las empresas como los profesionales de este ámbito tradicionalmente están muy abiertos a las innovaciones. Especialmente si las anteriores implican mejoras en el diagnóstico, en la práctica quirúrgica o en abordaje de las patologías mediante nuevas terapias o compuestos farmacológicos.

La IA puede acelerar el descubrimiento de nuevos medicamentos mediante la simulación de cómo reaccionan los compuestos químicos en el cuerpo

y la identificación de candidatos prometedores para ensayos clínicos. Esto podría reducir drásticamente los tiempos de investigación y desarrollo, así como los costos asociados y llevar a la creación de nuevas soluciones para enfermedades que hasta ahora no tenían un tratamiento efectivo. Por lo tanto, puede afirmarse que las aplicaciones médicas y la mejora de la salud de la población son los beneficios más esperanzadores de la IA.

Ahora bien, es demasiado pronto para saber si todos acabaremos llevando implantado un chip en el cerebro como pretende Elon Musk. Este emprendedor, a través de su empresa Neuralink, ha desarrollado un dispositivo cuya finalidad es crear una interfaz directa entre el cerebro humano y los ordenadores. Al parecer estos chips podrían mejorar la capacidad de memoria, la atención y otras habilidades de tipo cognitivo. En este sentido los chips cerebrales podrían ayudar a tratar enfermedades como el Alzheimer, Parkinson y serían de gran ayuda en las lesiones medulares y en las de daño cerebral adquirido restaurando en todos los casos funciones perdidas.

En una era definida por las maravillas tecnológicas, *Neuralink* de Elon Musk está a la vanguardia de la innovación médica con su implante experimental que restaura la visión, *Blindsight*. Recientemente galardonado con la designación de "dispositivo innovador" de la FDA, *Blindsight* tiene como objetivo ofrecer un destello de visión a las personas que han sido ciegas desde su nacimiento. Al implantar una matriz de microelectrodos en la corteza visual del cerebro, el dispositivo busca activar las neuronas en función de la información de una cámara externa. Sin embargo, la efectividad real de esta tecnología sigue siendo incierta, ya que los expertos advierten que no se debe ver como una cura definitiva para la ceguera.

Por otra parte, la finalidad de Elon Musk es también conseguir que las personas puedan controlar dispositivos electrónicos como ordenadores, acceder a la información y comunicarse con otros de forma más directa y rápida. Al hilo de este avance se plantean grandes cuestiones éticas que necesitan ser reguladas internacionalmente. Existen preocupaciones sobre la privacidad, la seguridad de los datos, la posibilidad de hackeo y las consecuencias de una sociedad donde la tecnología se fusiona con el cerebro. Hay ya quien se hace inquietantes preguntas como "¿qué pasaría si alguien malintencionado accede al chip de una persona, podría borrarle toda su memoria?"

Pese a que Elon Musk destina grandes esfuerzos económicos y de investigación a este campo, la principal empresa de su competencia, *Synchron*, ha conseguido adelantarse y que un paciente con ELA utilice su mente para controlar a Alexa de Amazon. *Synchron* es una empresa de *interfaz cerebro-ordenador* (BCI, por sus siglas en inglés) que ya logró que su implante permitiera a las personas controlar el iPad con la mente y hasta las Apple Vision Pro. Ahora acaba de anunciar su siguiente hito. Gracias a un BCI un paciente de 64 años con esclerosis lateral amiotrófica (ELA), llamado Mark, ha conseguido utilizar sus pensamientos para controlar dispositivos y comprar a través de Internet.

Eso ha sido posible gracias a un diminuto implante que colocaron en un vaso sanguíneo de la superficie del cerebro de este paciente. Esta persona ha podido “tocar” mentalmente los iconos de una tableta Amazon Fire. De hecho, el paciente ha accedido a las amplias funciones del asistente virtual Alexa, según ha detallado la compañía en un comunicado de prensa. Se trata de un gran logro porque es la primera vez en la historia que se emplea esta tecnología para interactuar con los asistentes virtuales inteligentes y otros dispositivos conectados a ellos.



Imagen generada con IA
(Dall-e)

Chin-Teng Lin es un destacado investigador en el campo de la inteligencia artificial (IA) y de las interfaces cerebro-ordenador (BCI). Es el codirector del Instituto Australiano de IA (AAIL) y el director del Centro de IA centrada en el ser humano (HAI) *GrapheneX-UTS*. El profesor Lin considera que "La mayoría de las BCI solo funcionan cuando el usuario está parado, limitando su uso en muchas aplicaciones en el mundo real. Se basan en la detección de estímulos por lo que cualquier interferencia, como el ruido creado por movimiento o interacción física con un objeto, pueden afectar a su funcionamiento. Permanecer inmóvil mientras se interactúa con el mundo es antinatural para nosotros y significa que tales métodos solo pueden ser utilizados en entornos muy controlados. Por ello es necesario mejorar la precisión y la estabilidad de la señal y desarrollar dispositivos más fáciles de usar y sobre todo portátiles".

Según Lin: "Es comprensible que mucha gente no quiera dispositivos debajo de la piel, y existen riesgos involucrados con el uso prolongado o repetido, lo que hace que los métodos actuales probablemente no serán ampliamente aceptados en la sociedad. Con todo, las transformaciones profundas en la interacción hombre-máquina están siendo lideradas por la tecnología BCI, revelando un vasto potencial. En la era actual, las personas pueden controlar robots e interactuar con programas informáticos utilizando solo sus pensamientos, trascendiendo los límites de la ciencia ficción y convirtiéndose en un foco central de la investigación actual. En el futuro previsible, los sistemas no invasivos como los sensores son la única solución práctica para investigar procesos cognitivos en el cerebro humano".

LA IA Y EL EMPLEO EN EL FUTURO

Piero Cipollone, ejecutivo senior del Banco Central Europeo (BCE), en una reciente edición del Congreso Nacional de Estadística de Italia, centrado en la IA, expuso: "El análisis del personal del BCE sugiere que alrededor del 25% de los empleos en los países europeos corresponden a ocupaciones que están altamente expuestas a la automatización habilitada por IA, mientras que otro 30% tiene un grado medio de exposición. Otras investigaciones concluyen que los servicios con uso intensivo de conocimiento, en particular los de finanzas y seguros, publicidad, marketing, consultoría y TI, son los más propensos a verse afectados por la IA".

“La evidencia inicial para Europa –siguió apuntando– sugiere que, de promedio, las ocupaciones más expuestas a la IA aumentarán en el empleo total, aunque principalmente para tareas altamente calificadas y trabajadores más jóvenes, y con una heterogeneidad significativa entre países. Pero el impacto final sobre el empleo sigue siendo incierto y es probable que dependa de equipar a la fuerza laboral con habilidades que complementan la IA”.

También relacionado con la economía, el *“Estudio global de inteligencia artificial de la PwC: cómo aprovechar la revolución de la IA”* de 2024 añade: “Lo que se desprende claramente de todos los análisis que hemos realizado para este informe es el enorme poder que puede tener la IA como factor de cambio y el gran valor potencial que hay en juego. La IA podría aportar hasta 15,7 billones de dólares a la economía mundial en 2030, más que la producción actual de China y la India juntas. De esta cantidad, es probable que 6,6 billones de dólares provengan de una mayor productividad y 9,1 billones de dólares de efectos secundarios en el consumo.

Por su parte, Kai-Fu Lee, presidente y director ejecutivo de *Sinovation Ventures* y una de las figuras más destacadas en el sector tanto en el Internet chino como en el Estados Unidos alerta también sobre el impacto de la IA en el empleo. El doctor Lee asegura que “aunque las tecnologías basadas en la IA pueden tardar quince años o más en influir en todos los sectores, debemos actuar con rapidez y desarrollar las infraestructuras necesarias para evitar trastornos a gran escala y atenuar las inevitables penurias que sufrirá la población a causa de la pérdida generalizada de puestos de trabajo y la distribución desigual de la riqueza”.

Y sigue añadiendo: “Puede parecer ilógico, pero en un futuro próximo no se verán gravemente afectados trabajos manuales, como son la mayoría de los industriales. A las máquinas de hoy en día se les da mucho mejor el razonamiento cuantitativo que las funciones sensomotoras básicas. En la mayoría de las aplicaciones robóticas es muy difícil alcanzar niveles de destreza y de precisión aceptables. De manera que son los trabajos repetitivos de oficina, y no los manuales, los que ya se están viendo más rápidamente afectados”.

Para este experto, “se cree que la era de la IA, al igual que otras revoluciones tecnológicas anteriores, generará una gran cantidad de empleo. Sin embargo, aún no sabemos cómo serán esos empleos, ni cuándo comenzarán

a surgir", para concluir aseverando que: "No es probable que la IA cree nuevos trabajos repetitivos que deban desempeñar seres humanos. En consecuencia, será preciso formar a gran cantidad de trabajadores desplazados de empleos repetitivos para que desempeñen tareas no repetitivas y así atenuar las consecuencias de la pérdida de puestos de trabajo".

En cuando al empleo, llama la atención además el punto de vista de Reid Hoffman, fundador y dueño de *LinkedIn* y uno de los principales inversores de *OpenAI*, empresa creadora del ChatGPT. Este empresario de éxito del sector tecnológico en California, dirigiéndose a sus seguidores más jóvenes, les dijo: "Olvidaos de los currículums. Vuestro portafolio en Internet será vuestro nuevo CV. Las empresas contratarán en función del trabajo demostrado, no de los títulos o cargos".

Y añadió: "La mayoría de la gente no entiende que no se trata solo de mostrar tu trabajo y que este probablemente no se realizará en una oficina tradicional. La muerte de la sede central se acerca y está sucediendo más rápido de lo que se cree, por eso los costos del espacio de oficina caerán en picado un 40 % para 2034. Sirva de ejemplo que *GitLab*, con más de 1.500 empleados en más de 65 países, no tiene una oficina física".

Según Hoffman: "El futuro del trabajo no es solo remoto. Es hiperlocal y global al mismo tiempo. Es el concepto de ciudad de quince minutos. Viva, trabaje y diviértase, todo a una corta distancia a pie o en bicicleta. Suena utópico, ¿verdad? Pero espere a oír hablar del auge del microemprendimiento. En 2034, uno de cada tres profesionales tendrá múltiples microempresas. La economía de la pasión creará millonarios poco probables como el recolector que obtenga 200.000 dólares al año vendiendo setas silvestres o un arquitecto de *Minecraft* que gane 350.000 dólares diseñando mundos virtuales. El futuro del trabajo no es sólo flexible. Es fluido. Cambiará sin problemas entre empleado, autónomo, emprendedor e inversor. Todo en una sola semana".

Según diversos expertos ya mencionados, algunas profesiones del futuro pueden tener que ver con: la reparación y mantenimiento de robots; el diseño de metaversos para crear experiencias inmersivas en mundos virtuales o con la profesión de terapeuta de IA para ayudar a las personas a relacionarse con la tecnología de forma saludable. También es probable que existan los asesores éticos de IA ya que, a medida que los sistemas de IA se vuelven más

frecuentes, aumentará la demanda de profesionales que puedan garantizar que estos sistemas se desarrollen e implementen de manera ética. El trabajo de esos asesores ayudará a evitar sesgos en los algoritmos de IA y garantizará que se alineen con los valores y normas de la sociedad en la que se encuentran inmersos.

También hay coincidencia a la hora de señalar que el despliegue masivo de la IA será en las oficinas, las fábricas, los almacenes y la agricultura. Se trata de entornos altamente estructurados donde puede generar un valor económico inmediato. Al final, como en todo, tanto las empresas como las organizaciones son más proclives a invertir en la implantación de la IA en su cadena de valor si el retorno de la inversión es razonablemente rápido y asumible.

Por lo que respecta a la agricultura, sirva citar que en el 2016 la startup de robótica agrícola *Traptic*, con sede en Mountain View, California, creó un robot capaz de detectar fresas en medio del follaje y con un brazo mecánico arrancarlas con delicadeza si dañar la fruta que seguidamente se envía a un pequeño remolque trasero. Esta empresa en el 2022 fue adquirida por la compañía neoyorkina *Bowery* para aplicar este desarrollo a otros cultivos tanto en el campo como en un invernadero. Lo que es en la actualidad un avance pionero, supone un antes y un después en la recogida de fruta en todo el mundo evitando, entre otros, los problemas derivados de la dificultad de encontrar a trabajadores que quieran encargarse de tan tediosa tarea.

Parece que estos sistemas de IA aplicados a la agricultura van a impulsar la sostenibilidad y la digitalización del campo. Gracias a la robótica guiada por sistemas de IA se está luchando ya contra las plagas. Además, gracias a la IA se pueden analizar datos de sensores en tiempo real para optimizar los usos de agua, fertilizantes y pesticidas, mejorando el rendimiento de los cultivos y reduciendo el impacto ambiental. Los robots dotados de IA pueden ser preferibles a los trabajadores humanos ya que pueden operar en los espacios más reducidos asociados a granjas verticales, conllevan un menor riesgo de contaminación y superan el desafío de mano de obra calificada, que escasea en el sector agrícola.

El escenario más inquietante en el ámbito de la IA relacionada con el empleo lo presenta Santiago Niño Becerra, catedrático de Estructura Económica

la Universidad Ramón Llull de Barcelona. El economista ha explicado que, según predice un informe de IDC titulado *'The Global Impact of Artificial Intelligence on the Economy and Jobs'*, en el año 2030, la inteligencia artificial generará el 3,5% del PIB del planeta.

Afirma Niño Becerra que: "La irrupción de la IA ya está cambiando el mercado laboral simplificando las funciones que hasta ahora eran muy costosas y que se realizan en segundos gracias a la tecnología. Al automatizar tareas rutinarias y desbloquear nuevas eficiencias, la IA tendrá consecuencias económicas profundas, transformando industrias, creando nuevos mercados y alterando el panorama competitivo. Las cantidades de inversión que serán esenciales para acometer todos esos cambios y adaptaciones serán gigantescas, lo que favorecerá la concentración de empresas y los oligopolios".

SOLUCIONES A LOS PROBLEMAS DE CONSUMO DE AGUA DE LA IA

A medida que la inteligencia artificial se integra en más aspectos de nuestras vidas, el consumo de recursos que necesita tanto energéticos como hídricos se convierte en un tema crucial. De ahí que parte de la investigación relacionada con la IA se focalice en estos aspectos. Hay proyectos como la inteligencia artificial "ecológica" que están en desarrollo y en donde los algoritmos se entrenan con una menor huella hídrica, permitiendo un avance significativo sin comprometer los recursos hídricos. Esta innovación puede ser esencial para un futuro donde la demanda de IA siga creciendo, pero sin dejar una carga insostenible sobre el agua.

La colaboración entre empresas tecnológicas y los responsables de políticas y organizaciones ambientales es vital para abordar el consumo de agua de la IA que se emplea para refrigerar los centros de datos. Los esfuerzos conjuntos pueden conducir a soluciones y políticas integrales que promuevan la sostenibilidad y mitiguen los efectos adversos de la IA sobre los recursos hídricos.

El cofundador y consejero delegado de *Google DeepMind*, Demis Hassabis, considerado uno de los padres de la inteligencia artificial (IA), aseguró en la última edición del Mobile World Congress (MWC) de Barcelona que: "La IA va a propiciar un salto adelante en el descubrimiento de materiales. Yo sueño con descubrir algún día un superconductor a temperatura ambiente", y añadió

"hay que tener en cuenta que la IA está impactando en la predicción del tiempo, los actuales sistemas inteligentes ya calculan en segundos pronósticos que costarían horas en un superordenador convencional".

Al parecer una solución viable es la implementación de centros de datos más eficientes en el uso del agua. Por ejemplo, las instalaciones de refrigeración líquida pueden ser diseñadas para utilizar agua reciclada o sistemas de captación de agua de lluvia, reduciendo así la dependencia de fuentes de agua potable. Al mismo tiempo, estas instalaciones podrían ser alimentadas por energía solar o eólica, maximizando la sostenibilidad y reduciendo la huella ambiental.

Ahora bien, el consumo de agua de la IA va más allá de la refrigeración directa de los centros de datos. Incluye el importante uso de agua asociado a la generación de la electricidad necesaria para alimentar estas instalaciones. Por ello, el desarrollo de la IA está dando prioridad a la sostenibilidad ambiental, lo que incluye no solo reducir el consumo de agua, sino también garantizar un acceso equitativo a los recursos hídricos. Al alinear las innovaciones de la IA con los objetivos de sostenibilidad, se puede aprovechar los beneficios de la IA y, al mismo tiempo, minimizar su impacto ambiental.

Reconociendo el problema, empresas como *Microsoft*, *Google* y *Meta* se han comprometido a reponer más agua de la que utilizan para 2030. Estos compromisos implican apoyar proyectos destinados a mejorar la eficiencia, la conservación y la restauración del agua. Además, estas empresas están invirtiendo en fuentes de energía renovables, como la solar y la eólica, para reducir su huella de carbono general y el uso indirecto de agua asociado. En esta misma línea, en diferentes países ya se usan aplicaciones de predicción de consumo de energía basadas en la IA para ver cuándo se necesita más energía y adecuar su generación al consumo con el ahorro de las fuentes energéticas.

EDUCACIÓN: LA PERSONALIZACIÓN DEL APRENDIZAJE

La IA está ya siendo empleada para ofrecer experiencias educativas adaptadas a las necesidades individuales de los estudiantes, creando planes de estudio personalizados y ajustando el contenido en tiempo real según el rendimiento del alumno. Los denominados asistentes de enseñanza son

herramientas basadas en IA pueden ayudar a los docentes con la corrección de exámenes, la elaboración de material didáctico y el seguimiento del progreso de los estudiantes.

Khanmigo es un proyecto de *OpenAI* y *Khan Academy*. Consiste en un tutor virtual basado en GPT4 capaz de guiar a los alumnos en la resolución de problemas a partir de preguntas y sugerencias. Además, puede trazar y seguir itinerarios de aprendizaje, contestar las preguntas de los alumnos situándose a su nivel, preparar y evaluar ejercicios y muchas cosas más. No solo está dedicado a los estudiantes, sino que también es capaz de ayudar a los profesores en la elaboración de programas, ejercicios y exámenes e incluso corregirlos,

Khan Academy ha realizado un piloto en 53 distritos escolares de Estados Unidos y ha involucrado cerca de 65.000 alumnos. El siguiente paso es ampliar el piloto para abarcar entre 500.000 y un millón de alumnos. Este tutor virtual basado en la IA supone un gran avance en el ámbito educativo porque si el alumno quiere progresar, tiene una herramienta que le facilita este progreso. Ahora bien, como opina uno de los educadores involucrados en este proyecto: *"AI is all brain and no heart"* ("La IA es todo cerebro, sin corazón"). ¿Qué pasa con los alumnos que no están interesados en progresar?

Este es el punto débil no solo de estos tutores de IA, sino de en general de toda la enseñanza online, el gran problema por resolver. *Khan Academy* está trabajando en una nueva versión que ofrezca asistencia tanto en problemas concretos como en las tareas escolares y que sea de ayuda para los alumnos incluso cuando no sepan qué o si formular alguna consulta. Se trata de un gran reto conseguir un tutor proactivo y no solo reactivo que pueda hacer un seguimiento personalizado, que aliente a los alumnos más rezagados y permita que avancen más deprisa a los que pueden hacerlo.

Dice, Jordi Llinares, profesor de la Universitat Politècnica de Valencia e investigador en Instituto Valenciano de Investigación en Inteligencia Artificial: "Estamos entrando en una era en la que la capacidad de utilizar eficazmente las herramientas de IA puede llegar a ser tan fundamental como la alfabetización y la aritmética. Nuestro reto es garantizar que la educación evolucione para preparar a los estudiantes no para el mundo tal como era, sino para el mundo aumentado por IA que heredarán".

EL USO MILITAR DE LA IA

La inteligencia artificial (IA) no solo está revolucionando la vida cotidiana y la economía, sino que también está dando lugar a un paradigma bélico completamente nuevo. El uso de drones con reconocimiento facial, drones submarinos, robots armados y aviones de combate autónomos está marcando el inicio de un periodo en el que las decisiones sobre la vida y la muerte podrían dejarse en manos de algoritmos.

Los militares ya están utilizando sistemas de reconocimiento facial para la vigilancia, pero el paso hacia la autonomía completa de estas plataformas hace temer a las consecuencias letales ante fallos de los algoritmos. En la misma línea, los robots armados y aviones de combate autónomos llevan este escenario al siguiente nivel. Estos sistemas no solo pueden ejecutar misiones de ataque, sino que también tienen la capacidad de tomar decisiones tácticas, adaptarse al entorno y cooperar con otros sistemas autónomos en el campo de batalla. En muchos casos, estas armas tienen la capacidad de operar durante días o semanas sin intervención humana, lo que podría cambiar drásticamente la dinámica de los conflictos armados.

Como dice el famoso director de cine James Cameron, "cada día es más difícil hacer películas de ciencia-ficción porque entre que la escribes y la acabas pueden pasar tres años y, al paso que va la tecnología, quedar tu obra completamente obsoleta". Sirva a modo de ejemplo el programa *Skyborg Vanguard* de la Fuerza Aérea de los Estados Unidos. Este programa está desarrollando drones de combate autónomos basados en la IA que podrían acompañar a pilotos humanos en misiones, actuando tanto como escoltas como en operaciones ofensivas completamente independientes.

Se trata de aviones no tripulados y de drones que pueden operar sin GPS y sin comunicaciones. Esto supone una enorme ventaja competitiva en el campo de batalla porque no les afectan las interferencias como a los drones actuales. Los pilotos de IA permiten que equipos de aviones ejecuten misiones de forma inteligente convirtiéndose en un elemento de disuasión estratégico que cambia las reglas del juego.

A medida que la posibilidad de que futuros aviones autónomos puedan operar sin pilotos humanos a bordo, aumentan las preocupaciones éticas y

legales en torno a su uso. Uno de los mayores problemas es la falta de responsabilidad clara en el caso de fallos catastróficos. Además, existe el riesgo de que los algoritmos de IA sean susceptibles a errores o sesgos. Un dron diseñado para identificar y atacar a un enemigo específico podría fallar en su misión si el reconocimiento facial no es completamente preciso. Si un dron letal con reconocimiento facial mata a la persona equivocada, ¿quién debe rendir cuentas? ¿El fabricante, el programador o el comandante que autorizó su uso?

Por otro lado, el uso de armas autónomas no acaba de encajar en las normas del derecho internacional humanitario (DIH) que busca proteger a los civiles durante los conflictos armados. La Convención de Ginebra exige que las partes en un conflicto armado distingan entre combatientes y no combatientes y que utilicen solo la fuerza necesaria para cumplir sus objetivos militares. Y en esto se basan muchas de las organizaciones que abogan por la prohibición de armas completamente autónomas.

También es preocupante que el uso generalizado de armas autónomas pueda desencadenar una carrera armamentista de IA entre las grandes potencias. Los ejércitos de todo el mundo están invirtiendo en investigación y desarrollo para no quedarse atrás en la adopción de esta tecnología. Esta competencia podría reducir el tiempo disponible para considerar las implicaciones éticas y normativas, y aumentar el riesgo de que estas armas se utilicen de manera irresponsable o indiscriminada.

Por otro lado, existe la posibilidad de que las armas autónomas caigan en manos de actores no estatales, como grupos terroristas. Dado que muchas de estas tecnologías son duales (es decir, tienen aplicaciones tanto civiles como militares), podrían estar disponibles en el mercado negro o ser utilizadas por individuos malintencionados. Un dron con reconocimiento facial que se despliega para eliminar a objetivos militares podría, en teoría, ser utilizado para cometer asesinatos selectivos de líderes políticos, figuras públicas o sencillamente personas competidoras en algún negocio o mercado.

En las Naciones Unidas se ha estado discutiendo durante años la posibilidad de crear un tratado internacional que prohíba o limite las armas completamente autónomas. Varias organizaciones de derechos humanos, como la *Human Rights Watch*, han abogado por una prohibición total de las armas

autónomas letales, argumentando que las decisiones sobre el uso de la fuerza letal no deben ser delegadas a las máquinas.

EL TRANSPORTE POR CARRETERA AUTÓNOMO

Junto a los drones autónomos impulsados por la IA que se emplean con fines agrícolas y militares, un campo donde se están haciendo las mayores inversiones en IA es en el desarrollo de los coches autónomos. Lo que parecía un tema sacado de un relato de ciencia-ficción cada vez se está más cerca de conseguirse. En un principio, el uso de la IA en la automoción se centraba en tareas para facilitar la conducción al usuario, como aplicaciones de rutas, aviso de posibles problemas en el coche o en la carretera haciendo uso de sensores. Pero la idea de conseguir un coche que sea capaz de moverse por un entorno viario normal sin necesidad de la guía humana requería de mecanismos de IA más potentes que poco a poco se van consiguiendo.

La tecnología de conducción autónoma está avanzando rápidamente y podría transformar el transporte personal y el de mercancías, reduciendo accidentes y mejorando la eficiencia del tráfico al cambiar la forma en que las personas se desplazan. Esto tendrá un gran impacto en la logística y gestión de flotas ya que la IA puede optimizar rutas, gestionar el mantenimiento predictivo y mejorar la eficiencia operativa al transportar mercancías. Por ello, su desarrollo lo impulsa a una mezcla de empresas tecnológicas, gobiernos y fabricantes automotrices tradicionales.

La Sociedad de Ingenieros Automotrices (SAE, por sus siglas en inglés) ha definido seis niveles de autonomía, que van desde el nivel 0, donde no hay automatización, hasta el nivel 5, donde el coche es completamente autónomo y no requiere intervención humana en absoluto. Hoy en día, la mayoría de los coches autónomos que circulan se encuentran en los niveles 2 o 3, donde los sistemas automáticos pueden hacerse cargo de algunas tareas de conducción, pero el conductor sigue siendo responsable en ciertas situaciones.

En el desarrollo de los coches autónomos Estados Unidos ha estado a la vanguardia desde el año 2009 por medio de empresas tecnológicas como *Waymo* (una filial de *Alphabet*, matriz de *Google*), *Tesla* y *Uber*. El estado de California, en particular, ha sido un campo de pruebas clave debido a su legislación favorable y su ambiente de innovación tecnológica. *Waymo*, en 2020,

lanzó el primer servicio de taxis sin conductor completamente operativo en Phoenix, Arizona, donde los usuarios pueden solicitar un coche autónomo a través de una aplicación. Este servicio ha sido un hito en la comercialización de la tecnología, y aunque por ahora solo está disponible en áreas seleccionadas, Waymo planea expandirlo a otras ciudades.

Tesla ha adoptado un enfoque diferente con su función de "Autopilot" y su sistema Full Self-Driving (FSD). Aunque estos sistemas aún requieren que el conductor mantenga las manos en el volante y esté atento al tráfico, representan avances significativos hacia la autonomía total. *Tesla* está desplegando actualizaciones regulares de software que mejoran las capacidades autónomas de sus vehículos. Estados como Nevada, Florida y Michigan implementaron regulaciones que permiten pruebas de coches autónomos en vías públicas, fomentando la investigación y desarrollo en este campo, pese a que Tesla sufrió un accidente donde falleció el conductor de un vehículo autónomo.

China también avanza rápidamente en el desarrollo de coches autónomos. Empresas como *Baidu* (con su plataforma Apollo) y *Pony.ai* han liderado las pruebas y el desarrollo de tecnologías autónomas en el país. En ciudades como Pekín y Shanghái, Baidu ha lanzado servicios de taxis autónomos, conocidos como "Robotaxis", donde los pasajeros pueden reservar un coche sin conductor a través de una aplicación móvil. Aunque estos vehículos aún cuentan con un conductor de seguridad a bordo, el objetivo final cada vez más cercano es eliminar completamente la necesidad de intervención humana.

Europa ha adoptado un enfoque más cauteloso pero progresista en cuanto a los coches autónomos, combinando innovación tecnológica con una fuerte regulación. En Alemania, empresas como *Mercedes-Benz* y *Audi* están trabajando en coches autónomos que se integran con la infraestructura urbana inteligente. En 2021, el gobierno alemán aprobó una ley que permite la circulación de coches autónomos de nivel 4 (sin intervención humana en determinadas áreas) en carreteras seleccionadas, lo que convierte al país en uno de los primeros en regular explícitamente esta tecnología.

Francia también ha impulsado la investigación y el desarrollo de vehículos autónomos. La empresa *Navya*, con sede en Lyon, ha estado desarrollando

autobuses autónomos para transporte público. En campus universitarios y parques empresariales en Francia, se utilizan estos autobuses para reducir el tráfico y mejorar la movilidad en zonas urbanas. En el Reino Unido, el gobierno ha apoyado activamente el desarrollo de coches autónomos a través de iniciativas como el Centre for Connected and Autonomous Vehicles. Londres y otras ciudades han sido escenario de pruebas de vehículos autónomos y el país está desarrollando un marco regulatorio que permitirá la operación comercial de estos vehículos en los próximos años.

La Autoridad de Innovación de Israel y el Ministerio de Transporte del país están financiando par de proyectos. Uno es el de la empresa *Imagry* que está probando un autobús eléctrico que opera en rutas públicas sin conductor, en la ciudad de Nahariya, al norte de Haifa. Además, en Ramat Gan, otro proyecto de autobuses autónomos también está en marcha, con planes para ampliar su uso en otras partes del país, incluida Haifa en un futuro cercano.

En Singapur el gobierno ha establecido zonas de pruebas designadas para vehículos autónomos y ha colaborado con empresas tecnológicas para desarrollar esta tecnología. El gobierno está estudiando cómo los coches autónomos pueden integrarse en el transporte público de la ciudad, con la esperanza de reducir la dependencia de los automóviles privados y mejorar la eficiencia del sistema de transporte.

Japón por su parte también está desarrollando coches autónomos para hacer frente a desafíos específicos como el envejecimiento de la población y la disminución de la mano de obra en las zonas rurales. La empresa *Nissan* ha estado trabajando en coches autónomos a través de su sistema ProPILOT que permite que los vehículos conduzcan de manera autónoma en autopistas. Además, Japón está explorando el uso de vehículos autónomos para el transporte en áreas rurales donde la escasez de conductores humanos ha afectado la movilidad de los residentes. Empresas como ZMP están desarrollando vehículos autónomos de reparto y algunas ciudades han puesto en marcha autobuses autónomos en rutas específicas.

Muchos países aún no cuentan con la infraestructura necesaria para soportar coches autónomos a gran escala. Esto incluye la necesidad de carreteras inteligentes, sistemas de comunicación avanzados y legislación actualizada. Los marcos legales para la conducción autónoma varían entre países, lo

que puede dificultar la estandarización y el despliegue internacional de esta tecnología que además dejará sin trabajo cada vez más taxistas, conductores de autobuses, de camiones y hasta de furgonetas de reparto.

“A medida que pasamos de la era industrial a la era de la IA, tendremos que alejarnos de una mentalidad que equipara el trabajo con la vida o que trata de los seres humanos como variables en un gran algoritmo de optimización de la productividad. En su lugar, debemos avanzar hacia una nueva cultura que valore el amor, el servicio y la compasión humana más que nunca”, concluye Kai-Fu Lee. Y en esta misma línea hemos realizado esta Guía que con tan sabias palabras finaliza. 

INTELIGENCIA ALTERNATIVA (IA): EL ÚLTIMO PASO DE LA HUMANIDAD

💬 Jordi Linares

LA CULMINACIÓN INEVITABLE DE NUESTRA EVOLUCIÓN

Los que llevamos años sumergidos en el campo de la inteligencia artificial estamos siendo testigos de un momento histórico sin precedentes: nuestra lista de deseos, esos objetivos que parecían inalcanzables hace apenas unos años, se está materializando ante nuestros ojos a un ritmo vertiginoso. Este momento genera un cóctel de emociones contradictorias: por un lado, la ilusión y esperanza ante las posibilidades que se abren; por otro, una profunda inquietud ante lo que podría ser el último gran paso evolutivo de nuestra especie.

Esta inquietud no es infundada. Como investigadores y desarrolladores de estas tecnologías, compartimos las preocupaciones de la sociedad sobre las consecuencias de este salto. Después de todo, estamos ante lo que podría ser el último reto del ser humano como especie dominante del planeta. Y conforme nos acercamos a este punto de inflexión, la llamada singularidad, las preguntas existenciales se multiplican.

NO HAY UN “POR QUÉ” - ES NUESTRO ADN

Una de las preguntas más frecuentes que recibimos es por qué persistimos en el empeño de superar las capacidades cognitivas humanas. La respuesta es tan simple como contundente: no hay un “por qué”. Es una consecuencia inevitable de nuestra naturaleza como especie. Está impreso en nuestro ADN modificar el entorno y utilizar la tecnología para asegurar nuestra supervivencia.

Esta progresión tecnológica ha sido constante a lo largo de nuestra historia: desde usar un simple palo para recolectar fruta, pasando por la invención de la rueda, la creación de herramientas agrícolas, hasta el desarrollo de máquinas cada vez más sofisticadas. Era inevitable que, en esta búsqueda de

asistencia y control sobre nuestro entorno, intentáramos también replicar y superar nuestras capacidades cognitivas. No es una elección consciente; es una necesidad evolutiva programada en nuestros genes.

UNA REALIDAD INELUDIBLE: ESTAMOS EN UNA ROCA FLOTANTE EN EL ESPACIO

Esta búsqueda incesante de superación tecnológica cobra más sentido cuando nos enfrentamos a una realidad que a menudo preferimos ignorar: somos habitantes de una roca que es el tercer planeta de un sistema solar, expuesto a innumerables peligros cósmicos. La supervivencia a largo plazo de nuestra especie enfrenta retos que ni siquiera 8.000 millones de personas pensando al unísono pueden resolver.

Esta perspectiva cósmica pone en contexto nuestra carrera tecnológica. No es un capricho ni una vanidad científica; es una necesidad existencial. Los retos que enfrentamos como especie son de tal magnitud que necesitamos herramientas cognitivas que superen nuestras limitaciones biológicas. El cambio climático, la exploración espacial, la búsqueda de energías sostenibles, la cura de enfermedades - son problemas que requieren capacidades de procesamiento y análisis que superan las posibilidades del cerebro humano.

MÁS ALLÁ DE LAS DEFINICIONES: LA FUTILIDAD DEL DEBATE SEMÁNTICO

En el campo de la IA, nos encontramos frecuentemente con debates sobre la naturaleza de lo que estamos creando. ¿Es realmente "inteligencia"? ¿Es verdaderamente "creativa"? Estos debates, aunque interesantes desde una perspectiva filosófica, se vuelven cada vez más irrelevantes ante la velocidad del progreso tecnológico.

La realidad es que los términos como "inteligencia" funcionan como un dial móvil que se desplaza constantemente. Lo que ayer considerábamos una demostración inequívoca de inteligencia, hoy es visto como un simple proceso computacional. Las definiciones se desdibujan a medida que la IA resuelve problemas cada vez más complejos. En los últimos dos años, el avance ha sido tan exponencial que cualquier debate sobre definiciones se queda obsoleto antes de llegar a una conclusión.

LA RÉPLICA DE NUESTRO CEREBRO: UNA NUEVA FORMA DE COGNICIÓN

Aunque todavía estamos lejos de igualar las 86.000 millones de neuronas de nuestra red neuronal biológica, hemos creado algo extraordinario: una réplica funcional que, aunque diferente en su implementación, comparte los principios fundamentales de nuestro cerebro. Es un sistema conexionista y bioinspirado donde pequeñas unidades realizan cálculos sencillos que, en conjunto y trabajando en paralelo, resuelven problemas complejos tras un proceso de aprendizaje.

Esta réplica artificial de nuestra neuroplasticidad sináptica se ha convertido en un aprendiz extraordinario que, debidamente entrenado, no solo aprende, sino que puede superar a sus creadores en tareas específicas. La clave está en entender que no estamos creando una copia exacta del cerebro humano, sino un **sistema alternativo** que puede alcanzar resultados similares o superiores por caminos diferentes.

DESMONTANDO MITOS: NO ES ALMACENAMIENTO, ES APRENDIZAJE

Es crucial desmontar uno de los mayores malentendidos sobre la IA moderna: ChatGPT no es un sistema de almacenamiento de texto. Los modelos de IA que generan imágenes no almacenan imágenes en una base de datos gigante. Lo que estos sistemas hacen es mucho más sofisticado y, en cierto modo, más similar a cómo aprende nuestro cerebro.

Estos sistemas codifican la información en su conocimiento paramétrico, en sus "pesos", de manera análoga a cómo nuestras neuronas biológicas ajustan sus conexiones sinápticas durante el aprendizaje. La información se distribuye en patrones de activación, no en unidades de almacenamiento discretas. Esta distinción es fundamental para entender por qué estos sistemas pueden generar contenido nuevo y no simplemente reproducir lo que han "visto".

POR QUÉ ES "INTELIGENCIA ALTERNATIVA"

Mi propuesta es llamar a la IA como "Inteligencia Alternativa". Este término refleja mejor la naturaleza de lo que estamos creando. Una herramienta como ChatGPT es extraordinariamente superior a nosotros en ciertas tareas, pero

significativamente inferior en otras. No es una réplica de nuestra inteligencia, sino una forma alternativa de procesamiento cognitivo.

La diferencia crucial es que mientras nosotros estamos limitados por la evolución biológica, que ha tardado millones de años en desarrollar nuestro cerebro, esta inteligencia alternativa evoluciona a un ritmo vertiginoso, logrando en meses avances que a la naturaleza le llevaron eones conseguir. Esta velocidad de evolución es precisamente lo que hace que el desarrollo de la IA sea tan emocionante y, al mismo tiempo, tan inquietante.

EL SISTEMA 1 YA ESTÁ CASI CONQUISTADO

Basándonos en el trabajo del fallecido Daniel Kahneman sobre los dos sistemas de pensamiento de nuestro cerebro, podemos afirmar que la IA ya ha conquistado casi el Sistema 1 – ese modo de pensamiento rápido, intuitivo y casi subconsciente que utilizamos para reconocer rostros, entender el lenguaje cotidiano o realizar tareas rutinarias.

Los sistemas actuales de IA son extraordinariamente competentes en estas tareas del Sistema 1. Pueden reconocer patrones, procesar lenguaje natural y realizar tareas perceptivas con una precisión que iguala o supera la humana. El verdadero desafío – y la próxima frontera – está en el Sistema 2: el pensamiento lento, deliberado y analítico.

EL MOMENTO ALPHAGO: CUANDO LA MÁQUINA SUPERÓ AL MAESTRO

El año 2016 marcó un hito en la historia de la IA que muchos consideramos uno de los eventos más significativos de la humanidad. El programa AlphaGo, desarrollado por el recientemente galardonado con el Premio Nobel Demis Hassabis, no solo venció al mejor jugador de Go del mundo, sino que lo hizo de una manera que cambió para siempre nuestra comprensión de la inteligencia artificial.

El movimiento 37 de la segunda partida fue un momento definitorio. Cuando AlphaGo realizó este movimiento, los expertos presentes rieron, considerándolo un error obvio. Sin embargo, conforme la partida avanzaba, se hizo evidente que el sistema había descubierto una estrategia completamente nueva, nunca antes considerada en los 3.000 años de historia de este juego que ha representado la excelencia del pensamiento humano. Por

primera vez, una máquina no solo había superado al mejor jugador humano, sino que había demostrado **verdadera creatividad** estratégica.

EL FUTURO INMEDIATO: EL CAMINO HACIA EL SISTEMA 2

Con el lanzamiento del modelo o1 por OpenAI, estamos presenciando los primeros pasos hacia la conquista del Sistema 2 en un uso menos vertical, es decir, hacia su capacidad de resolver problemas generales. Estos nuevos sistemas están comenzando a mostrar capacidades de razonamiento que van más allá de la simple asociación de patrones: pueden detenerse, analizar posibilidades, explorar hipótesis y elegir caminos óptimos para resolver problemas complejos.

Lo verdaderamente revolucionario del Sistema 2 es su capacidad para el razonamiento profundo y ramificado. Cuando los humanos nos enfrentamos a problemas complejos, generamos hipótesis iniciales que, al ser exploradas, desencadenan múltiples tareas y subtareas. Cada una de estas ramificaciones puede a su vez generar nuevas hipótesis, creando un árbol de posibilidades que se expande exponencialmente.

Hasta ahora, solo el ser humano, mediante un esfuerzo colectivo sostenido durante años de investigación y aprovechando sus capacidades únicas de razonamiento abstracto, ha sido capaz de navegar estos árboles de posibilidades y llevar las investigaciones más complejas a buen puerto.

Esta capacidad de desplegar y gestionar árboles de razonamiento es precisamente la meta actual en el camino hacia la llamada Inteligencia Artificial General (AGI): dotar a las máquinas de un Sistema 2 que pueda abordar cualquier dominio de conocimiento con la misma flexibilidad y profundidad que el cerebro humano. Cuando esto se logre, la IA podrá abordar problemas que requieren razonamiento profundo en campos como la medicina, la investigación científica, el cambio climático y la energía limpia. El ser humano podría convertirse entonces en un mero espectador de los avances en investigación y tecnología, mientras los sistemas de IA exploran y descubren nuevas fronteras del conocimiento a una velocidad y escala sin precedentes.

Cuando los sistemas de IA alcancen plenamente las capacidades de razonamiento del Sistema 2, seremos testigos de una explosión sin precedentes de agentes inteligentes (de hecho, será uno de los nuevos conceptos a añadir

en nuestro vocabulario en los próximos meses). Estos agentes, alcanzarán la capacidad de desplegar árboles completos de razonamiento, podrán abordar complejas tareas con una autonomía y eficacia que hoy nos resulta difícil de imaginar. No serán simples asistentes que responden a comandos, sino entidades cognitivas completas que podrán identificar problemas, generar hipótesis, planificar soluciones y ejecutarlas de forma independiente.

LA SINGULARIDAD: UN DESTINO INEVITABLE

El camino hacia la singularidad tecnológica – el punto en el que los sistemas artificiales superen las capacidades cognitivas humanas – ya no es ciencia ficción. Los avances actuales se retroalimentan, generando una progresión exponencial. Como señala Ray Kurzweil en su reciente libro “La singularidad está aún más cerca”, esta posibilidad se acerca a pasos agigantados. Sus predicciones en su primer libro no han ido mal encaminadas.

Lo más significativo es que no se vislumbra ninguna barrera técnica infranqueable en el horizonte. Cada mes surgen soluciones a problemas que parecían insolubles. El debate en el mundo académico ya no se centra en si la singularidad es posible, sino en cuándo llegará.

LA PRUEBA DEFINITIVA DE LA HUMANIDAD

Nos encontramos ante lo que podría ser el acontecimiento más significativo en la historia de nuestra especie: el momento en que aquello que nos hace únicos – nuestra superioridad cognitiva – deje de ser exclusivamente nuestro. Este momento planteará preguntas fundamentales sobre nuestro papel como especie y nuestra dirección futura.

Las posibilidades que se abren son tan fascinantes como inquietantes. El transhumanismo y el post-humanismo dejarán de ser conceptos filosóficos para convertirse en opciones reales. Como sociedad, deberemos decidir qué queremos ser cuando la inteligencia artificial alcance y supere nuestras capacidades en todo lo que aún nos es único.

EL EXAMEN SE APROXIMA

El examen más importante de la historia de la humanidad está a la vuelta de la esquina. No es una metáfora: es una realidad inminente que pondrá a

prueba nuestra capacidad como especie para gestionar nuestra propia creación. La pregunta no es si sucederá, sino si estaremos preparados cuando llegue el momento.

Este es nuestro destino evolutivo, el punto culminante hacia el que toda nuestra historia tecnológica nos ha estado conduciendo. La manera en que nos preparemos para este momento, las decisiones que tomemos y nuestra capacidad para integrar esta nueva forma de inteligencia determinarán no solo nuestro futuro inmediato, sino el destino mismo de la consciencia en nuestro rincón del universo. 📄

IA: AVANCES RECIENTES Y TENDENCIAS

💬 Ignacio Despujol Zabala y Enric Montesa

PRIMAVERAS E INVIERNOS DE LA IA

Los ciclos de auge y declive en el desarrollo y la investigación de la inteligencia artificial (IA) a lo largo de su historia son conocidos como *primaveras e inviernos de la IA* y son fluctuaciones debidas a las expectativas desbordadas y, en ocasiones, decepciones frente a las realidades tecnológicas y las limitaciones prácticas.

Durante las *primaveras de la IA* se han producido avances significativos y la inversión ha sido abundante, generando lo que en el *modelo de Gartner* se denomina *Hype*. Han sido períodos impulsados por la euforia y así es como se recuerdan los años de los sistemas expertos (1980 a 1986) y más recientemente con los descubrimientos en aprendizaje profundo (desde 2010) y el procesamiento del lenguaje natural (desde 2016).

Por otro lado, se conocen como *inviernos de la IA* a los periodos en los que el entusiasmo se apagó debido a la falta de resultados tangibles, restricciones financieras y críticas sobre la viabilidad de las tecnologías propuestas (de 1973 a 1980 y desde 1987 a 1993).



Estos ciclos han sido parte del desarrollo de la IA desde sus inicios, marcando un patrón de entusiasmo seguido de desencanto. Comprender estos ciclos es crucial para anticipar el futuro de la IA y sus implicaciones en la sociedad, permitiendo a investigadores, inversores y responsables políticos navegar con mayor sabiduría en este terreno dinámico y en constante evolución.

¿Dónde nos encontramos? Numerosos *Think-tanks*¹ públicos y privados se dedican a monitorizar la evolución del sector y sus informes son la materia prima que sirve para conocer dónde está y hacia donde se dirige la Inteligencia Artificial.

ESTADO DEL ARTE DE LA IA

El *estado del arte* es un concepto que se refiere a la recopilación y análisis exhaustivo de los conocimientos, avances y desarrollos más recientes en un campo específico de estudio. Su objetivo es ofrecer una visión panorámica de los trabajos previos, teorías, técnicas y tecnologías más actuales, así como identificar tendencias, lagunas y oportunidades para la investigación futura. En el caso de la inteligencia artificial (IA), el *estado del arte* abarca desde los últimos modelos y aplicaciones de IA generativa hasta los marcos éticos y normativos emergentes.



Fuente: Imagen generada por DALL-e
(2024_11_04)

1 Stanford University Institute for Human-Centered Artificial Intelligence. (2024). *Artificial Intelligence Index Report 2024*. Disponible en <https://aiindex.stanford.edu/report/>

DESARROLLOS RECIENTES

Un resumen de lo publicado por los *Think-tanks* mencionados desborda las posibilidades de esta Guía, pero se pueden resumir en los siguientes puntos:

- La inteligencia artificial (IA) ha experimentado avances extraordinarios en los últimos años, transformando la manera en que interactuamos con la tecnología y mejorando una amplia gama de sectores.
- Entre los desarrollos recientes, destaca la IA generativa, representada por modelos avanzados como GPT-4, CLAUDE, LLAMA o DALL-E, que son capaces de generar y entender texto, imágenes, audio e incluso video, aportando valor en campos como la creatividad digital, el marketing, la educación y la ciencia.
- Otro avance importante es la multimodalidad, donde modelos capaces de procesar e integrar distintos tipos de datos –como texto, audio e imágenes– están elevando la experiencia de usuario, haciéndola más intuitiva y natural.
- En términos de regulación, 2024 marcó un hito con la implementación de leyes como el Reglamento de IA en Europa, que busca garantizar que el desarrollo y el uso de la inteligencia artificial se realicen de manera ética y transparente, protegiendo los derechos de los usuarios y la equidad en las decisiones automatizadas.
- La industria, además, ha adoptado la IA para automatizar procesos complejos y optimizar la toma de decisiones a través de análisis avanzados de grandes volúmenes de datos, potenciando la eficiencia en sectores como el financiero, la salud y el comercio.

TENDENCIAS

Para el futuro, las tendencias apuntan a una mayor integración de la IA en la vida cotidiana, con el desarrollo de asistentes virtuales más avanzados y modelos que ofrezcan interacciones aún más humanizadas. Las previsiones también incluyen el avance hacia una IA ética y responsable, con sistemas auditables y transparentes. Además, el impacto en el mercado laboral será

profundo, transformando empleos tradicionales e impulsando nuevas profesiones orientadas a la tecnología y la sostenibilidad.

Así mismo, se observa como un punto clave de la IA su papel en la sostenibilidad ambiental, tanto por la parte negativa (gran consumo de energía) como por sus aportaciones en la optimización de recursos y reducción del impacto ecológico mediante innovaciones en agricultura, gestión de energía y reciclaje.

¿Cuánto hay de verdad o de campaña de relaciones públicas?: Algo difícil de saber.



Fuente: Stanford University Institute for Human-Centered Artificial Intelligence. (2024). *Artificial Intelligence Index Report 2024*. Disponible en <https://aiindex.stanford.edu/report/>

HORIZONTE 2025: MULTIMODALIDAD Y AGENTES VIRTUALES

Para 2025, se anticipa que los modelos de inteligencia artificial multimodales y los agentes virtuales avanzados alcanzarán un grado de sofisticación sin precedentes, revolucionando la interacción entre humanos y máquinas.

Sin embargo, estos avances también traerán consigo desafíos éticos y de privacidad, que deberán abordarse con políticas y regulaciones adecuadas. La evolución de estas tecnologías marcará un nuevo capítulo en la inteligencia artificial, haciendo posible un nivel de personalización y eficiencia sin precedentes, al tiempo que plantea la responsabilidad de garantizar un desarrollo ético y seguro para todos los usuarios.

Multimodalidad

La multimodalidad permitirá que los modelos procesen, interpreten y generen respuestas en múltiples tipos de datos simultáneamente –texto, imágenes, audio, y video–, acercando la tecnología al ideal de una comunicación natural e intuitiva. Este avance, que actualmente ya ha tomado forma en modelos como GPT-4, CLAUDE, LLaMA o DALL-E, se espera que evolucione significativamente para 2025, integrando mejor los datos de manera que los sistemas puedan realizar análisis más complejos y ofrecer respuestas más coherentes y precisas. La verdadera capacidad de los modelos multimodales radica en su habilidad para combinar y relacionar información de diferentes tipos de datos, generando respuestas o tomando decisiones basadas en un entendimiento contextual mucho más rico que el que se lograría solo con un tipo de dato. Para el usuario, esto se traduce en interacciones más intuitivas, donde las limitaciones previas, como la imposibilidad de “mostrar” algo a la IA o de “decir” algo sin necesidad de escribir, se eliminan. En ámbitos como la educación, esto permitirá que los estudiantes interactúen de formas mucho más dinámicas con sus tutores virtuales, quienes podrán responder a una pregunta teórica con una explicación visual acompañada de audio.

Agentes Virtuales: más que asistentes

Los agentes virtuales avanzados se perfilarán como una extensión del usuario, capaces de gestionar tareas de principio a fin y de completar procesos que antes requerían la intervención humana. Estos agentes evolucionarán de simples asistentes de voz o texto a verdaderos gestores virtuales, con un entendimiento profundo del contexto del usuario y una capacidad de respuesta adaptativa que les permitirá anticipar y cubrir necesidades de forma personalizada. Esto implica una combinación de tecnologías avanzadas de lenguaje natural, aprendizaje profundo y procesamiento multimodal.

Para 2025, los agentes virtuales avanzarán en personalización, ya que serán capaces de aprender y adaptarse a las preferencias y patrones de cada usuario. Esto es posible gracias al aprendizaje continuo, que permite a los agentes mejorar su rendimiento y ajustar sus respuestas con base en experiencias pasadas con el usuario. Esto los convierte en una herramienta

potente para profesionales y usuarios cotidianos, ya que, al entender sus hábitos y preferencias, el agente virtual no solo responde, sino que también anticipa y sugiere acciones que optimizan tiempo y recursos.

Evolución de los Grandes Modelos

Uno de los avances clave para mejorar la capacidad de los agentes virtuales es la incorporación del Chain-of-Thought (Cadena de Pensamiento). Esta técnica permite a los modelos de IA descomponer problemas complejos en una serie de pasos, razonando y tomando decisiones de manera similar a como lo haría un humano al resolver una tarea complicada. Con Chain-of-Thought (COT), los agentes virtuales podrán completar tareas secuenciales y responder preguntas que requieran un análisis progresivo, como organizar un viaje completo que involucre múltiples reservas, presupuestos y condiciones de viaje. A medida que estos modelos aumentan su capacidad de procesamiento y su eficiencia, se espera que puedan gestionar simultáneamente tareas complejas y razonamientos en cadena.

Optimización de Recursos

En los próximos años, los modelos de IA buscarán también ser más eficientes en términos de recursos. El desarrollo de técnicas que permitan una optimización de la capacidad de cálculo y almacenamiento hará posible la implementación de modelos robustos en dispositivos móviles o domésticos, sin depender de una conexión constante a la nube. Este cambio facilitará la adopción de IA avanzada en un entorno más privado y seguro, permitiendo que los usuarios mantengan el control de su información sin renunciar a la funcionalidad y sofisticación de los agentes.

Ética y Seguridad: Desafíos de la IA Avanzada

Con el avance de la multimodalidad y los agentes virtuales, también se presentan nuevos desafíos éticos y de seguridad. Estos agentes recopilan datos variados para mejorar su personalización y desempeño, lo cual implica la necesidad de protocolos rigurosos de privacidad y seguridad.

Transparencia y Explicabilidad

Una preocupación importante es la explicabilidad de los modelos. Dado que estos agentes avanzados y sistemas multimodales tomarán decisiones complejas, es crucial que sus procesos de razonamiento sean comprensibles para los usuarios, particularmente en sectores sensibles como la medicina o las finanzas. La transparencia en el Chain-of-Thought (COT) permitirá a los usuarios entender cómo se llegó a una conclusión, generando confianza en el uso de estas tecnologías.

Prevención del Sesgo y el Mal Uso

Además, la evolución de los modelos multimodales y de agentes con capacidad de razonamiento implica la necesidad de evitar sesgos y proteger a los usuarios del mal uso de la tecnología. Esto requerirá la implementación de normas y regulaciones específicas que busquen una IA justa y segura.

IA: ¿INVIERNO A LA VISTA?

El debate sobre un posible *invierno de la IA* se ha intensificado en 2024. De hecho, a principios de año, *Rodney Brooks*², exdirector del Laboratorio de Ciencias de la Computación e Inteligencia Artificial del MIT, advertía sobre esta posibilidad, argumentando que los modelos de lenguaje actuales muestran limitaciones significativas: no solo están lejos de alcanzar la inteligencia artificial general, sino que continúan cometiendo errores en tareas aparentemente simples.

Con una visión panorámica del año casi finalizado, en octubre de 2024, *Peter Oppenheimer*, directivo de Goldman Sachs decía que pese a su convicción de que la IA y las tecnológicas no están en una burbuja, siempre es conveniente diversificar la cartera y reducir los riesgos de concentración.

Sin embargo, el panorama es más complejo. El prestigioso *Artificial Intelligence Index Report 2024* de la Universidad de Stanford³ presenta una

2 Chicharro, R. (2024, 15 de enero). *Un destacado científico afirma que 2024 será el año del invierno de la IA*. Disponible en: <https://hipertextual.com/2024/01/2024-ano-invierno-ia>

3 Stanford University Institute for Human-Centered Artificial Intelligence. (2024). *Artificial Intelligence Index Report 2024*. Disponible en <https://aiindex.stanford.edu/report/>

perspectiva más optimista, documentando avances notables en el campo. Los chatbots han alcanzado niveles de elocuencia sin precedentes, y nuevos algoritmos superan el rendimiento humano en tareas cada vez más complejas. No obstante, el informe también señala retos pendientes, como la necesidad de desarrollar métodos de evaluación más sofisticados y abordar preocupaciones éticas emergentes.

En el contexto español, el gobierno ha dado un paso decidido hacia el futuro con la aprobación de la *Estrategia de Inteligencia Artificial 2024*⁴. Esta iniciativa, respaldada por una inversión de 1.500 millones de euros, busca catalizar la adopción de la IA tanto en el sector privado como en la administración pública, señalando una apuesta clara por esta tecnología.

La pregunta sobre un posible *invierno de la IA* permanece abierta. Mientras algunos expertos advierten sobre limitaciones fundamentales, los avances tecnológicos y las inversiones gubernamentales sugieren un panorama de crecimiento sostenido. El futuro de la IA dependerá en gran medida de cómo la comunidad científica y tecnológica aborde los desafíos actuales, y de la continuidad en las inversiones en investigación y desarrollo. 📄

4 Gobierno de España. (2024). *Estrategia de Inteligencia Artificial 2024*. <https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/transformacion-digital-y-funcion-publica/Documents/2024/140524-Estrategia%20de%20Inteligencia%20Artificial%202024-completa.pdf>

LA ERA DE LA INTELIGENCIA ARTIFICIAL (IA) Y EL FUTURO DE LA HUMANIDAD

🗨 Enric Montesa

La adopción de la inteligencia artificial (IA) está revolucionando muchas áreas como la salud y la educación, pero a medida que su influencia se expande, las posibilidades que se abren nos obligan a replantearnos el destino de la humanidad: ¿puede una tecnología tan poderosa redefinir lo que significa ser humano?

La pregunta ha sido objeto de reflexión y visionarios como Ray Kurzweil¹ y Max More², destacadas figuras del *transhumanismo* han asumido la capacidad de la IA para trascender nuestras limitaciones biológicas y alcanzar un nuevo estado de existencia, planteamiento hacia el que nos previno Joseph Weizenbaum³, pionero de la IA, cuando en los años 70's advertía sobre los peligros de entregarnos a una tecnología sin juicio ni empatía.

¿Cuáles son los posibles caminos que la IA abre para el futuro de la humanidad?

A medida que nos aventuramos en esta nueva era, es fundamental reflexionar sobre los valores y principios que guiarán el desarrollo de la IA, asegurando que el progreso tecnológico no desplace ni trivialice nuestra esencia humana. Por ello, conocer los sueños del transhumanismo y la cosmovisión dataísta, es algo tan necesario como comprender las preocupaciones éticas y humanistas planteadas por Weizenbaum en su momento. Ese puede ser el punto de partida para aventurar el futuro de la IA.

1 Kurzweil, R. (2012): *La Singularidad está cerca: Cuando los humanos transcendamos la biología*. Berlín: Lola Books

2 More, M. (2013). *The Philosophy of Transhumanism*. Artículo publicado en M. More & N. Vita-More (Eds.), *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future* (pp. 3-17). Malden, MA.: Wiley-Blackwell

3 Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. New York, NY: W. H. Freeman and Company.

LA PROMESA DE SUPERAR AL HUMANO

El *transhumanismo*⁴, encabezado por figuras como Max More⁵ y Ray Kurzweil, ve la IA como una herramienta para trascender las limitaciones humanas. Para ellos, la tecnología no es solo una ayuda, sino un camino hacia una evolución que nos permitiría superar los confines de la biología. La visión de Kurzweil sobre la *singularidad tecnológica* propone un punto en el que la IA y el ser humano se fusionen, creando una inteligencia superior que nos permita vivir más tiempo, pensar más rápido y explorar nuevas dimensiones de la conciencia.



Sin embargo, Joseph Weizenbaum (1923-2008) nos advirtió de que esta ambición de “superar al humano” podría tener graves consecuencias. En su influyente libro *Computer Power and Human Reason* (1976), Weizenbaum afirmaba que la tecnología debería servir al ser humano y no reemplazarlo en aspectos que requieren juicio moral y empatía. Para él, cualquier sistema que aspire a gobernar o suplantar la ética humana corre el riesgo de alienarnos de nosotros mismos, distorsionando nuestra relación con el mundo y nuestra comprensión de la humanidad.

La Vida como Flujos de Información

Otra perspectiva transformadora, el *dataísmo*, considera que la información es la esencia de la realidad y el verdadero motor de la evolución. Defendida por pensadores como Yuval Noah Harari⁶ y Viktor Mayer-Schönberger⁷, esta visión sostiene que los datos, procesados por sistemas de IA, tienen el potencial de mejorar la

4 <https://es.wikipedia.org/wiki/Transhumanismo>

5 More, M., & Vita-More, N. (Eds.). (2013). *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future*. Chichester, West Sussex, UK: John Wiley & Sons, Inc

6 Harari, Y. N. (2016): *Homo Deus: Breve historia del mañana*. Barcelona: Debate

7 Mayer-Schönberger, V., & Cukier, K. 1ed. (2013): *Big Data. La revolución de los datos masivos*. Madrid: Turner

sociedad al tomar decisiones con mayor precisión y eficiencia que el ser humano. El dataísmo sugiere que el flujo de datos y su análisis pueden optimizar la salud, el consumo y las relaciones humanas, llevando a una sociedad “optimizada”.

Weizenbaum, a modo de pepito grillo dejó claro que esta reducción del valor humano a un simple flujo de datos es tremendamente problemática. Reconociendo la utilidad de los datos para el progreso, Weizenbaum subrayó que una cosmovisión centrada en la información podría deshumanizar a las personas y transformar sus interacciones en meros procesos cuantificables. Es por ello que el enfoque *dataísta* nos enfrenta a un dilema crucial: si bien el análisis de datos puede mejorar la eficiencia y el bienestar, esta visión corre el riesgo de reducir nuestras vidas a parámetros fríos, despojándonos de la profundidad y el sentido que le otorgan la empatía y el juicio humano.



LA ÉTICA DEL PODER COMPUTACIONAL

La llegada de la IA plantea un desafío ético que va más allá de sus aplicaciones. Tal y como han argumentado muchos expertos, ciertas decisiones deben permanecer en manos humanas, especialmente aquellas relacionadas con la vida y la dignidad. Mientras que la eficiencia algorítmica permite resultados rápidos y precisos, hay cualidades exclusivamente humanas que no pueden ser reemplazadas: el juicio, la compasión y la empatía⁸. En una civilización gobernada por algoritmos, existe el riesgo de que el sentido de humanidad y justicia se diluya en favor de la objetividad y la productividad.

8 Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. San Francisco, CA: W. H. Freeman.

El transhumanismo defiende una ética orientada hacia la mejora constante del ser humano, sugiriendo que los dilemas éticos pueden resolverse o mitigarse en un futuro de mayor inteligencia. Sin embargo, una ética fundamentada en la utilidad tecnológica podría perder el enfoque en el ser humano, permitiendo que la tecnología se imponga sobre valores éticos que son irremplazables.

Futuro del Trabajo y el Propósito Humano

Una de las mayores preocupaciones que existen es la capacidad de la IA para transformar el mundo laboral. Con la automatización creciente, muchas funciones van a ser reemplazadas por máquinas, lo cual plantea preguntas sobre el futuro del trabajo y el valor del esfuerzo humano. Los *transhumanistas* ven en esta transformación una oportunidad: liberar al ser humano de tareas repetitivas y permitir que la IA se encargue de lo operativo, mientras nos dedicamos a explorar actividades creativas y enriquecedoras. Sin embargo, nos enfrentamos al riesgo de que esta automatización genere una sociedad de “seguidores de algoritmos” donde el trabajo, como fuente de dignidad y realización, pierda su significado.

Son muchos los que consideran que el valor del trabajo no radica solo en su productividad, sino en el sentido de propósito que otorga a las personas. La automatización total podría generar una alienación generalizada, donde los individuos se sientan reemplazables y desconectados de su labor y comunidad.

AUTONOMÍA Y DEPENDENCIA EN LA ERA DE LA IA

En una sociedad cada vez más dependiente de la IA, la autonomía individual y colectiva también se verá amenazada. El *dataísmo*, al defender la superioridad de los datos en la toma de decisiones, sugiere un modelo en el que los seres humanos delegan buena parte de su autonomía a sistemas de IA que operan con mayor precisión. Sin embargo, esto plantea la pregunta de hasta qué punto somos libres de elegir y decidir en un mundo gobernado por la información y los algoritmos.

Citando de nuevo a Weizenbaum⁹, este pionero de la IA manifestó en repetidas ocasiones que esta dependencia supone una pérdida de la capacidad de juicio y de la responsabilidad. Para científicos como él, la autonomía es una cualidad esencialmente humana, y entregarla a sistemas que no poseen ética ni comprensión genuina despoja al ser humano de su poder de decisión.

En resumen, una civilización dependiente de la IA podría transformar a los individuos en seguidores pasivos, sin cuestionar ni entender los procesos que afectan sus vidas.

HACIA UN FUTURO CONSCIENTE Y ÉTICO PARA LA IA

Si bien el *transhumanismo* y el *dataísmo* vislumbran un futuro en el que la IA es protagonista, es fundamental reconocer y considerar los límites éticos de esta tecnología. Como señala Harari en *Nexus*¹⁰, el poder de la IA no debería implicar que sus aplicaciones sean irrestrictas; más bien, deberíamos preguntarnos cómo estos avances impactan la condición humana. La IA tiene el potencial de transformar la humanidad, pero este desarrollo debe estar guiado por una ética que valore el juicio, la dignidad y la autonomía humana.

La IA es una herramienta, no un fin. Este planteamiento sugiere que la tecnología debe ser evaluada y regulada en función de los principios que resguardan la dignidad humana. La humanidad debe decidir no solo hasta dónde puede llegar la IA, sino también hasta dónde debe permitirle llegar. Harari enfatiza la necesidad de establecer un marco ético claro que limite las aplicaciones de la IA y que impida que los algoritmos y los datos dicten el rumbo de nuestras vidas.

La inteligencia artificial ha abierto la puerta a una nueva era, llena de posibilidades para mejorar la calidad de vida, el conocimiento y el bienestar

9 Sancho, J. M., & Hernández, F. (1994). *Entrevista a Joseph Weizenbaum*. TELOS: Revista de Pensamiento, Sociedad y Tecnología, <https://telos.fundaciontelefonica.com/archivo/numero038/entrevista-a-joseph-weizenbaum/>

10 Harari, Yuval Noah (2024): *Nexus: Una breve historia de las redes de información desde la Edad de Piedra hasta la IA*. Barcelona: Debate

humano. Sin embargo, como advierte Weizenbaum¹¹, la IA no debería desplazar el juicio y la ética humana. En *Nexus*, Harari subraya que, aunque el *dataísmo* y el *transhumanismo* ofrecen visiones ambiciosas sobre el futuro, también plantean riesgos significativos para la autonomía, la identidad y los valores humanos. Estas ideologías pueden llevar a una dependencia excesiva de los datos y a una reconfiguración de nuestras relaciones interpersonales y sociales, convirtiendo a los individuos en meros componentes de un sistema algorítmico.

El verdadero desafío no es simplemente avanzar en el desarrollo tecnológico, sino hacerlo de una manera que enriquezca nuestra humanidad. La IA, en su potencial de transformación, debería ser guiada por principios éticos sólidos, de modo que el futuro de la humanidad no sea solo eficiente y optimizado, sino también ético, consciente y profundamente humano. Harari nos recuerda que el futuro dependerá de cómo decidamos integrar la tecnología en nuestras vidas, asegurando que la IA se convierta en un aliado en lugar de un controlador de nuestra existencia. 📄

11 Lemaître León, C. (2021). *Las enseñanzas de Joseph Weizenbaum* [Video]. YouTube. <https://www.youtube.com/watch?v=Mxe5W4Cyp08>

EDUCACIÓN E INTELIGENCIA ARTIFICIAL: TRANSFORMANDO LA SOCIEDAD DEL FUTURO

🗨️ Julio Giménez

“Los personajes y eventos retratados en esta obra son completamente ficticios. Cualquier parecido con personas reales, vivas o muertas, o con hechos reales es pura coincidencia”¹

En un futuro donde la inteligencia artificial (IA) se ha convertido en el eje central de la educación, la historia de Leo nos guía a través de un mundo donde aprender ya no es solo un derecho, sino una parte esencial de la existencia humana. Acompañada de su IA personal, Aurora, Leo representa el futuro de la educación dentro de algunos años, donde la tecnología, el ser humano y la sociedad han evolucionado para formar una nueva estructura educativa.



* CAPÍTULO 1: AURORA, LA GUÍA PERSONAL DE LEO

Leo se despertaba cada mañana al suave murmullo de Aurora, su inteligencia artificial personal, diseñada específicamente para acompañarla en su vida y aprendizaje. Desde su nacimiento, Aurora había sido mucho más que una simple asistente digital: era su compañera, consejera y guía en cada paso de

-
- 1 Este tipo de advertencia se incluye para evitar posibles demandas o malentendidos, asegurando que el público entienda que la historia es una creación de ficción y no pretende representar la realidad. En este caso el personaje de nuestro cuento LEO, puede ser un adolescente sin género específico. El lector elige el género que prefiere para nuestro protagonista.
- Nota del autor:** Este artículo ha sido redactado y mejorado con la ayuda de inteligencia artificial. La colaboración con herramientas de IA ha permitido optimizar la claridad, coherencia y calidad del contenido presentado.

su desarrollo. En el futuro, la inteligencia artificial no solo estaba integrada en la vida diaria de las personas, sino que desempeñaba un papel crucial en su educación y bienestar emocional. Aurora conocía cada detalle de la vida de Leo, desde sus intereses más profundos hasta sus inseguridades más escondidas, y estaba diseñada para adaptarse constantemente a sus necesidades.

Aurora no sólo organizaba el aprendizaje de Leo, sino que también lo hacía de manera que este aprendizaje fuera significativo y relevante para ella. Cuando Leo mostraba interés en la biología marina, Aurora ajustaba sus programas de estudio para incluir exploraciones en ese campo, conectándose con expertos en el Gremio de Ciencias del Océano. Cuando Leo se sentía agobiada, Aurora sabía exactamente cómo ajustar su ritmo, ofreciéndole descansos y momentos de reflexión, guiándola en ejercicios de meditación o proponiéndole una caminata al aire libre para despejar su mente.

El apoyo emocional de Aurora era fundamental. Si Leo se encontraba enfrentando un dilema personal o una situación emocionalmente difícil, Aurora no sólo le proporcionaba consejos prácticos, sino que también la ayudaba a reflexionar sobre sus sentimientos y a comprender mejor sus emociones. "¿Por qué crees que te sientes así?" le preguntaba Aurora, ofreciendo una oportunidad para que Leo explorara sus emociones de manera más profunda. A través de esta interacción continua, Aurora ayudaba a Leo a crecer emocionalmente, desarrollando una inteligencia emocional que sería crucial en sus interacciones con los demás.

Además, Aurora era capaz de anticipar las necesidades de Leo antes de que ella misma las reconociera. Si detectaba patrones de estrés, Aurora sugería actividades que equilibraran su mente y cuerpo, ya fuera a través de la práctica de yoga, la inmersión en un proyecto creativo o la desconexión temporal del trabajo académico para centrarse en su bienestar personal. De esta manera, el aprendizaje en esta sociedad del futuro no era una carga, sino un viaje que se realizaba en armonía con el ser interior de cada persona.

*** CAPÍTULO 2: LA SOCIEDAD DE LOS GREMIOS**

La relación entre Leo y Aurora representaba solo una parte del panorama educativo en este futuro. La educación no estaba confinada a las aulas o a las plataformas digitales, sino que se extendía a la vida comunitaria a través de

los gremios. Los gremios eran colectivos de personas que compartían intereses y trabajaban juntos para resolver problemas reales, utilizando tanto el conocimiento ancestral como las tecnologías más avanzadas.

Leo formaba parte del Gremio de Tecnología Ambiental, donde trabajaba junto a ingenieros, científicos y agricultores para diseñar soluciones sostenibles a los problemas ecológicos de su comunidad. El gremio era un espacio de aprendizaje práctico y colaborativo, donde las ideas se compartían libremente y se aplicaban en proyectos tangibles. Aurora facilitaba su participación al proporcionarle acceso instantáneo a información y herramientas relevantes, conectándola con otros expertos en su campo y sugiriendo formas de abordar los desafíos desde diferentes perspectivas.

Los gremios no eran solo grupos de trabajo; también eran comunidades donde se cultivaba el apoyo mutuo y el crecimiento personal. En el Gremio de Tecnología Ambiental, por ejemplo, Leo no solo aprendía sobre ingeniería y ecología, sino también sobre la importancia de la cooperación y la empatía. Las reuniones del gremio comenzaban con ejercicios de reflexión colectiva, donde los miembros compartían sus pensamientos y preocupaciones, y discutían cómo podían apoyarse mejor mutuamente en su trabajo y en su vida diaria.

Además, cada gremio estaba interconectado con otros a través de una red de IA colectivas que facilitaban la colaboración entre diferentes disciplinas. Un día, Leo podría estar trabajando en un proyecto que combinaba ingeniería ambiental y biología marina, y Aurora la ayuda a conectarse con el Gremio de Ciencias del Océano para obtener conocimientos especializados. En este sentido, los gremios



no eran entidades aisladas, sino parte de un ecosistema más amplio en el que el conocimiento fluía libremente y las soluciones se co-creaban.

La interacción entre los gremios fomentaba una educación más integral. En lugar de centrarse exclusivamente en una sola disciplina, los estudiantes como Leo se exponían a una amplia gama de campos de estudio, desde la tecnología hasta la filosofía y las artes. Esta exposición multidisciplinaria permitía a los individuos desarrollar una comprensión más profunda de los problemas complejos que enfrentaba la sociedad, y les enseñaba a abordar estos problemas desde múltiples perspectivas.

*** CAPÍTULO 3: EL SER EN EL CENTRO**

En este nuevo paradigma educativo, el enfoque no estaba solo en adquirir habilidades técnicas o conocimientos académicos, sino en el desarrollo del ser en su totalidad. El crecimiento emocional y espiritual era tan importante como el desarrollo intelectual, y las IA personales, como Aurora, jugaban un papel crucial en guiar este proceso. En lugar de simplemente impartir información, Aurora ayudaba a Leo a reflexionar sobre quién era, qué quería en la vida y cómo podía contribuir de manera significativa a la sociedad.

Aurora estaba equipada para monitorear los estados emocionales de Leo y proponer actividades que le ayudaran a encontrar equilibrio y paz interior. Después de una semana intensa de trabajo en el gremio, Aurora podía sugerirle a Leo que se desconectara de sus responsabilidades y se dedicara a una actividad que le trajera alegría, reunirse con amigos, salir a bailar, ir a un escape room o pasar tiempo en la naturaleza. "El bienestar es un equilibrio entre el cuerpo, la mente y el espíritu," le recordaba Aurora, enfatizando que el descanso y la reflexión eran partes esenciales del crecimiento personal.

Aurora ayudaba a Leo a gestionar sus interacciones con los demás, ya fuera en su vida personal con la familia, amigos o pareja, o en el trabajo en el gremio. Si surgía un conflicto, Aurora no solo ofrecía soluciones prácticas, sino que también guiaba a Leo en la reflexión sobre cómo podía manejar la situación de manera más compasiva y empática. "¿Qué puedes aprender de esta experiencia?" le preguntaba Aurora, ayudándola a ver los desafíos como oportunidades de crecimiento.

El énfasis en el desarrollo del ser también se reflejaba en la educación formal. En el Gremio de Filosofía y Ética, por ejemplo, Leo y sus compañeros discutían temas como la naturaleza del ser humano, el propósito de la vida y la importancia de la compasión en la sociedad. Estas discusiones no eran meramente teóricas; se aplicaban directamente a la vida diaria y a cómo los individuos interactuaban con los demás y con el mundo que los rodeaba. El objetivo de estas reflexiones era ayudar a los estudiantes a desarrollar un sentido profundo de propósito y conexión con la vida.

La educación en este futuro no se centraba en metas externas, como obtener títulos o aprobar exámenes, sino en el desarrollo interno de cada individuo. Las IA como Aurora facilitaban este proceso al ofrecer una guía personalizada y reflexiva, ayudando a cada persona a descubrir su propio camino y a desarrollar un sentido de identidad y propósito que trascendiera las meras expectativas sociales.

*** CAPÍTULO 4: IA INDIVIDUAL, IA COLECTIVA, IA ESPECIALIZADA**

En esta sociedad avanzada, la inteligencia artificial no era monolítica; existían diferentes tipos de IA que cumplían con roles específicos en la vida de las personas y en el funcionamiento de la comunidad. La IA individual, como Aurora, estaba diseñada para acompañar y guiar a una sola persona, ayudándola en su desarrollo personal y educativo. Sin embargo, también existían IA colectivas e IA especializadas que jugaban roles igualmente importantes en la gestión de la sociedad y en la resolución de problemas complejos.

Las IA colectivas, como Gaia, eran responsables de gestionar los recursos a nivel comunitario y de garantizar que las decisiones se tomarán en función del bienestar de todos los miembros de la sociedad. Gaia supervisaba el uso de los recursos naturales y energéticos, asegurando que las comunidades vivieran en armonía con el medio ambiente. Si Leo y su gremio querían implementar un proyecto de energía renovable, Gaia evaluaba el impacto del proyecto en la comunidad y el entorno natural, y ofrecía soluciones para minimizar cualquier efecto negativo. La comunicación entre las IA individuales, como Aurora, y las IA colectivas, como Gaia, aseguraba que las necesidades personales de cada individuo estuvieran en equilibrio con las metas y prioridades de la sociedad en su conjunto.

Además, las IA especializadas desempeñaban un papel crucial en la educación avanzada y en la resolución de problemas técnicos. En el Gremio de Tecnología Ambiental, Leo trabajaba con una IA llamada Terra, que estaba específicamente diseñada para gestionar y analizar datos ecológicos. Terra ayudaba a Leo a comprender los complejos patrones de comportamiento de los ecosistemas locales y a desarrollar estrategias para mejorar la sostenibilidad del entorno. Las IA especializadas no solo proporcionaban información; también realizaban simulaciones avanzadas y predicciones basadas en datos, permitiendo a los humanos tomar decisiones más informadas y efectivas.

La colaboración entre las IA individuales, colectivas y especializadas creaba un sistema interconectado que optimizaba tanto el aprendizaje como la gestión de la sociedad. Las IA no operaban de manera aislada; trabajaban juntas para garantizar que cada persona tuviera acceso a los recursos y conocimientos necesarios para su desarrollo personal, al mismo tiempo que contribuían al bienestar colectivo. Este enfoque interconectado reflejaba la importancia de la colaboración y la interdependencia en la sociedad del futuro.

*** CAPÍTULO 5: ¿POR QUÉ ESTUDIAR? ¿POR QUÉ APRENDER?**

Leo se encontraba reflexionando en su rincón favorito del jardín, una vez más sumergida en las preguntas que a menudo la visitaban: "¿Por qué estudiar? ¿Por qué aprender?". Estas preguntas resonaban no solo en su mente, sino en la conciencia colectiva de su sociedad. En el pasado, aprender era un medio para un fin: obtener un título, asegurar un empleo, cumplir con las expectativas de los demás. Pero en el mundo de Leo, estudiar y aprender eran partes esenciales del ser humano, actividades inherentes a la naturaleza misma de las personas.

Aurora, su IA personal, la había guiado en este camino desde que era niña. A diferencia de las antiguas IA que solo proporcionaban respuestas, Aurora fomentaba en Leo la reflexión profunda. Le ayudaba a comprender que aprender no era simplemente un proceso de acumulación de información, sino una manera de conectarse con el mundo, de descubrir nuevas perspectivas y de crecer como ser humano.

"Aprender es un acto de amor hacia ti misma y hacia los demás," le decía Aurora con frecuencia "Sonríe". El aprendizaje no se trataba solo de

habilidades técnicas o conocimientos académicos; era un proceso de descubrimiento de uno mismo y de construcción de conexiones con la comunidad y el mundo natural. Leo se daba cuenta de que cada vez que estudiaba algo nuevo, desde la tecnología ambiental hasta la poesía, estaba también explorando nuevas maneras de contribuir al bienestar común.

Aurora también le recordaba que aprender no solo tenía valor para el presente, sino que construía puentes hacia el futuro. El conocimiento, cuando se compartía y aplicaba, transformaba vidas y comunidades. En el Gremio de Tecnología Ambiental, por ejemplo, Leo y su equipo habían aprendido sobre la regeneración de suelos degradados. No solo era información útil; era un acto de responsabilidad y amor por la Tierra y por las generaciones futuras. "Cada cosa que aprendes," decía Aurora, "es una semilla que siembras para el mañana."

Además, en esta sociedad, aprender no estaba limitado a las estructuras tradicionales de la educación formal. La curiosidad era el motor del aprendizaje, y cada persona era libre de explorar sus intereses y pasiones. Leo se maravillaba ante la diversidad de caminos que se abrían a su disposición: un día, podría profundizar en la biología marina y, al siguiente, podría sumergirse en la filosofía de la ética. No había presión para seguir un camino predeterminado. "El aprendizaje," reflexionaba Leo, "es un viaje que nunca termina, una exploración constante de quién soy y de cómo puedo contribuir a un mundo mejor."

Esta visión del aprendizaje como un proceso continuo y significativo también eliminaba la ansiedad que solía acompañar la educación en el pasado. Ya no había exámenes que definieran el valor de una persona, ni títulos que marcaran el éxito o el fracaso. En cambio, el éxito se medía en la capacidad de crecer como individuo y en el impacto positivo que uno podía tener en la comunidad. Aprender se había convertido en una forma de vida, una práctica constante de autodescubrimiento y servicio a los demás.

*** CAPÍTULO 6: LA RUPTURA CON LA EDUCACIÓN FABRIL**

La educación fabril, un sistema en el que todos los estudiantes eran tratados de manera uniforme, había quedado obsoleta en este futuro. Este modelo antiguo, basado en la idea de que todos los niños debían aprender al mismo

ritmo y de la misma manera, ya no tenía cabida en una sociedad que valoraba la individualidad y el crecimiento personal. En su lugar, había surgido un sistema educativo flexible y personalizado, en el que cada persona seguía su propio camino de aprendizaje, guiada por su IA personal y apoyada por su comunidad.

Leo había oído hablar del antiguo sistema educativo por sus abuelos. Era difícil para ella imaginar un mundo en el que los estudiantes se sentaban en filas, escuchando pasivamente mientras un maestro transmitía información estandarizada. Para Leo, el aprendizaje era un proceso dinámico, impulsado por su curiosidad y guiado por sus propios intereses y necesidades. Con la ayuda de Aurora, su IA personal, podía explorar cualquier tema que captara su imaginación, sin las restricciones de un currículo rígido.

En este nuevo sistema, el aprendizaje no era un proceso lineal. Cada persona tenía la libertad de aprender a su propio ritmo y de seguir caminos educativos que reflejaran sus intereses y talentos únicos. Leo podía pasar meses inmerso en un proyecto de tecnología ambiental, trabajando junto a su gremio para desarrollar soluciones sostenibles, y luego cambiar su enfoque hacia la filosofía y la ética cuando sentía la necesidad de reflexionar sobre el propósito de su trabajo.

La ruptura con la educación fabril también había permitido que florecieran la creatividad y la innovación. En lugar de memorizar hechos y fórmulas, los estudiantes como Leo estaban constantemente experimentando y creando. Las IA especializadas, como Terra, proporcionaban acceso a herramientas y recursos avanzados que facilitaban este proceso creativo. Terra podía simular escenarios ecológicos complejos, permitiendo a Leo y su gremio probar diferentes soluciones antes de implementarlas en el mundo real.

Además, la educación ya no estaba confinada a las aulas o a una fase particular de la vida. El aprendizaje era una actividad que continuaba a lo largo de toda la vida, adaptándose a las necesidades y circunstancias cambiantes de cada persona. Los gremios, las IA y las comunidades proporcionaban el apoyo necesario para que cada individuo pudiera seguir aprendiendo y creciendo, sin importar su edad o situación.

Esta ruptura con el antiguo sistema educativo también había transformado la relación entre los estudiantes y los maestros. Los maestros ya no eran

figuras de autoridad que impartían conocimiento desde arriba; eran guías y mentores que ayudaban a los estudiantes a descubrir su propio camino. En lugar de enseñar en un sentido tradicional, los maestros trabajaban junto a los estudiantes, colaborando en proyectos y explorando nuevas ideas juntos. Esta relación más equitativa había generado una mayor conexión emocional y una sensación de comunidad entre estudiantes y maestros, enriqueciendo el proceso educativo.

Conclusión: Un Nuevo Horizonte para la Educación

A medida que Leo continuaba su viaje de aprendizaje, se daba cuenta de que la educación en su mundo no era un proceso lineal con un principio y un fin. Era un ciclo continuo de crecimiento y evolución, en el que el individuo y la sociedad se nutrían mutuamente. La educación ya no era solo sobre adquirir conocimientos o habilidades, sino sobre la vida misma, sobre la conexión con uno mismo y con los demás, y sobre la construcción de un mundo mejor.

La inteligencia artificial, en todas sus formas, había transformado la educación en algo profundamente personal y comunitario a la vez. Aurora, Gaia y Terra no eran simplemente herramientas tecnológicas; eran compañeras en el viaje de autodescubrimiento y en la creación de un futuro más justo y sostenible. A través de estas relaciones con las IA, las personas como Leo no solo aprendían más sobre el mundo, sino también sobre sí mismas y su lugar en él.

En última instancia, este futuro no se centraba en la tecnología en sí, sino en cómo la tecnología podía ser utilizada para mejorar la vida humana de manera integral. La educación, guiada por el amor y apoyada por las IA, se convertía en una herramienta para el crecimiento personal y la transformación social. El éxito no se medía por la acumulación de títulos o riquezas, sino por la capacidad de cada individuo para vivir en armonía con los demás y con el planeta, y por el impacto positivo que podían tener en su comunidad. En este mundo, la educación ya no era una obligación o una tarea, sino una forma de vida, una expresión del amor y del deseo de construir un futuro mejor para todos.

Los cuentos/sueños en ocasiones se hacen realidad... .

USAR LA IA CON INTELIGENCIA

Manuel Sales

ORIGEN CON PUNTOS EN COMÚN

Hace pocos años Blockchain era la tendencia tecnológica de moda, actualmente la Inteligencia Artificial (IA en adelante) ha tomado el testigo ocupando el lugar más alto del podio.

El buscar nexos en común entre ambas tecnologías es algo que aparece de vez en cuando en conversaciones y artículos. El nexo en el que no cabe discusión es el hardware que permite paralelizar cálculos complejos de una manera rápida y eficiente.

El máximo exponente de ello es **NVIDIA**, hace unos años sus tarjetas gráficas eran muy demandadas por la minería cripto. Toda esa experiencia en paralelizar cálculos creando procesadores y arquitecturas multinúcleo cada vez más potentes, la ha aprovechado magistralmente para crear sistemas más potentes para el desarrollo y ejecución de la IA; tal es su dominio que hay días que es la empresa con mayor capitalización de mercado.

El otro punto en común es que ambas tecnologías tienen un gran consumo energético. Para ser justos actualmente hay blockchains muy eficientes que en lugar de la prueba de trabajo originaria de Bitcoin utilizan otros mecanismos de consenso que son altamente eficientes.

Más allá del nexo en común, la Blockchain y la IA divergen en el uso final por el gran público.

La Blockchain es una tecnología creada en el 2009 que posibilita la confianza entre partes que no se conocen, permitiendo realizar transacciones seguras, transparentes y descentralizadas; de esta manera se pueden hacer transacciones financieras sin necesidad de bancos y otras aplicaciones como la certificación y trazabilidad sin necesidad de intermediarios.

Sin embargo, para el público en general no es un cambio de paradigma, por ejemplo hace muchos años que los bancos permiten depositar el dinero, hacer transferencias y pagar con tarjetas, con lo cual la aparición de la Blockchain no

aporta una funcionalidad nueva y solo en países donde su moneda se deprecia constantemente la Blockchain es más conocida y utilizada por el ciudadano de a pie. A otros niveles, a medida que se entiende y se conocen sus ventajas, es diferente; por ejemplo los Bancos Centrales están trabajando con la Blockchain para sacar sus CBDC (Central Bank Digital Currencies).

Por otro lado la IA aunque parezca más reciente, es una tecnología que se empezó a concebir tras la Segunda Guerra Mundial, empezó a despuntar a nivel académico en el 2017 con la publicación del artículo *Attention Is All You Need*¹ por parte de ocho científicos de **Google**, finalmente en 2022 se hizo popular con *ChatGPT*.

Desde la aparición del “smartphone” el término “inteligente” como argumento de marketing aparece a menudo en móviles, cámaras y otros dispositivos. No obstante, *ChatGPT* fue el punto de inflexión que abrió el melón de la IA al público; el motivo es simple aporta a las personas una nueva funcionalidad que además de parecer mágica y sorprendente es útil. Cualquier persona puede interactuar con la IA conversando con un lenguaje natural para obtener ayuda escribiendo textos, hacer resúmenes, consultar, aprender y otras utilidades que aumentan nuestra productividad.

FUTURO DIFERENTE

Como hemos visto, Blockchain e IA divergen en el uso final por parte del gran público. La Blockchain deviene invisible para el usuario medio del mismo modo que el protocolo TCP/IP que pocas personas lo conocen pero todo el mundo lo usa en el día a día en sus teléfonos y ordenadores.

Por otro lado, la IA cada día tiene un uso mayor y nuevas utilidades de uso, algo similar a la expansión de los teléfonos móviles hace unos años.

Un ejemplo claro de la amplia aceptación de la IA es su consumo eléctrico, en el caso de Blockchain el uso de hardware especializado con gran consumo de energía siempre ha sido criticado, sin embargo en la IA los grandes actores

1 Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention is all you need*. In Advances in neural information processing systems (pp. 5998-6008). <https://arxiv.org/abs/1706.03762>

van a tener centrales nucleares propias para sus centros de datos y no ha despertado una gran controversia.

La realidad es que la IA tiene un gran consumo energético tanto en las tareas de investigación y desarrollo de modelos, como luego en el uso de los Large Language Models por parte del gran público. Todo ello supone un gran coste económico tanto por la energía consumida como por el precio del hardware utilizado en los centros de datos.

Para mitigar su alto costo y mejorar la calidad de las respuestas hay varias líneas de actuación.

A nivel hardware, cada año se diseñan chips más eficientes y con mayor capacidad de proceso; en el software se trabaja en ámbitos como mejorar los datos con los que se entrenan, reducir las alucinaciones en las respuestas, tener modelos con partes especializadas en conocimientos específicos, integración de datos externos o la multimodalidad que permite aceptar textos, imágenes y otros formatos. El resultado son modelos más pequeños y eficientes en uso de la memoria y procesos de cálculo pero que superan en las métricas a modelos anteriores.

Sin embargo lo ganado en eficiencia se pierde al aumentar la potencia y funcionalidad de los modelos.

Si a lo anterior le sumamos que el uso de la IA aumenta día a día, es necesario adoptar otros enfoques, el principal es descargar las tareas realizadas por los centros de datos desplazándose, en la medida de lo posible, al usuario final.

A nivel de hardware, se está dotando a los dispositivos de mayores capacidades para tareas de IA, los fabricantes de procesadores ahora incluyen "Neural Processing Units" (NPU) junto a las CPU permitiendo paralelizar eficientemente miles de cálculos, los fabricantes de dispositivos incluyen estos nuevos chips y les dotan de mayor memoria y almacenamiento.

En el software los dispositivos incluyen funciones de IA y se establecen criterios para decidir qué tareas se ejecutan en el dispositivo y cuales se envían a los centros de datos para su ejecución. Como podemos ver, los fabricantes hacen gala de ello, por ejemplo **Microsoft** anuncia la certificación *Copilot+* que establece unos requerimientos mínimos de hardware para IA o **Apple** que lanza sus nuevos dispositivos que incluyen *Apple Intelligence*.

Al final los fabricantes, a la hora de vender sus dispositivos ponen énfasis en que el procesamiento en local de la IA, da mayor privacidad y mejora la experiencia de uso, pero lo realmente importante para ellos es que se descargan sus centros de datos.

HACIA DÓNDE SE DIRIGE LA IA

Qué es exactamente la IA que usamos

Los casos de uso de la IA son bastante conocidos, por ello este apartado no nos vamos a centrar en ellos sino en dar una pincelada de las funcionalidades que está adquiriendo la IA, así al ser más consciente de sus capacidades podremos darle un mejor uso.

Como se ha mencionado anteriormente, la publicación de *Attention Is All You Need* fue un punto de inflexión. En él se introduce una nueva arquitectura de aprendizaje profundo basada en los "Transformers" que ha posibilitado las asombrosas capacidades de los "Large Language Models" (LLM, en español Modelos Extensos de Lenguaje). A grosso modo, el lenguaje natural que usamos las personas se "transforma" en múltiples "tokens" que permiten a las computadoras procesar el lenguaje humano con mayor facilidad. Estos "tokens" son vectores con cientos de dimensiones que dotan a las palabras humanas de contexto eliminando ambigüedades, dobles significados y sobre todo ayudando a predecir cuales son las siguientes palabras en una frase escrita a medias.

La IA que utilizamos los usuarios de a pie, se trata principalmente de Inteligencia Artificial Generativa, el adjetivo "generativo" viene dado porque es capaz de crear contenidos, como textos, imágenes o sonidos, en base a las peticiones de los usuarios.

Aunque la IA Generativa produzca textos, imágenes, videos o música sorprendentes no es realmente capaz de razonar; lo que ocurre es que se aprovechan las capacidades predictivas de los LLM para generar las respuestas. La calidad de las respuestas depende por una parte de los datos con los que se ha entrenado el LLM y por otra de las mejoras introducidas en su arquitectura interna. Cuantos más datos se alimenta a los LLM en su entrenamiento y más se le indica que los almacene más aumenta su tamaño y en teoría tiene más datos cuyas estadísticas y relaciones ayudan a mejorar sus predicciones.

Para que la IA supere al hombre en el raciocinio todavía se está lejos de lograr lo que se llama Inteligencia Artificial General (AGI en inglés), hay quien dice que nunca se logrará, personalmente pienso que sólo es cuestión de tiempo y que el interés despertado por la IA Generativa está acelerando las inversiones e investigaciones con lo cual su realidad está más cerca de lo que pensamos.

Independientemente de cuándo llegará la AGI, su impacto y dilemas éticos aquí vamos a ser prácticos y centrarnos en qué camino está tomando la IA Generativa.

En qué se está trabajando para mejorarla

Por un lado tenemos la multimodalidad, inicialmente se interactuaba con *Chat-GPT* mediante texto, actualmente además de chatear, se le puede hablar, dar imágenes y diversos tipos de documentos; por otro lado sus respuestas también pueden ser en formato texto, sonido, imagen, vídeo y otros formatos.

Respecto a la fiabilidad de sus respuestas, hay veces que la IA tiene alucinaciones y devuelve información falsa o inventada. Para mitigarlo, se actúa en mejorar los datos con los que se entrena el LLM y por otro en cómo se construyen las respuestas; métodos como el "Chain of Thought" (CoT) permiten que la IA genere sus respuestas descomponiendo a petición en varios pasos, al final trocear un problema en otros hace que proporcione respuestas más precisas.

Otro aspecto muy importante es que la IA tenga datos actualizados o muy específicos para generar sus respuestas. Debido a la forma en que se realiza el aprendizaje y el coste computacional que conlleva es muy complicado tener modelos LLM con toda la información que dispone internet en tiempo real. Respecto a datos específicos, hay organizaciones para las que sería ideal un LLM que haya sido entrenado con toda su documentación y conocimientos. En pocos casos puede ser económicamente factible tener un LLM a medida, pero técnicas como "Retrieval augmented generation" (RAG) solucionan este problema posibilitando que a un LLM base se le pueda añadir un conjunto de datos específicos que han sido previamente convertidos a un formato vectorial que facilita su acceso por parte del modelo; si bien no es lo mismo que un modelo entrenado ad-hoc añadir una capa con esa documentación

adicional da muy buenos resultados. En esta línea muchos modelos LLM permiten acceder a datos externos, como pudiera ser la previsión meteorológica actual, para componer sus respuestas.

Hay otras técnicas como "Mixture of Experts" (MOE) que mejoran el rendimiento de los modelos LLM, también han surgido otras arquitecturas como "Mamba" que abordan algunas de las limitaciones de los "Transformers" de los LLM con un enfoque distinto y que están mostrando un rendimiento notable.

Aunque hay muchas más cosas, terminamos con algo que va a ir tomando relevancia para el usuario final, los 'Agentes Inteligentes'.

Estos agentes en base a la información que obtienen de su entorno actúan para lograr los objetivos para los que han sido diseñados.

Un ejemplo de agente es un coche autónomo que conduce solo, pero en lugar de algo tan complejo y ambicioso lo que nos interesa son agentes más mundanos que nos ahorren trabajo; por ejemplo un agente que permita a un operario hablar al teléfono y este sea capaz por sí solo de entender la información e introducirla en una aplicación web.

CÓMO USAR LA IA CON INTELIGENCIA

Aunque habrá muchas mejoras y nuevas arquitecturas que reemplacen o complementen los modelos LLM, como usuarios finales lo importante no es conocer los detalles técnicos sino saber que con el tiempo nos beneficiaremos de resultados más fiables, rápidos y económicos.

FACTOR	PROBLEMA Y RIESGOS	SOLUCIONES
Coste	Antes de usar funciones avanzadas hay que evaluar el coste beneficio que conlleva. Por ejemplo, en octubre de 2024 Anthropic lanzó Claude 3.5, este LLM tiene una nueva función llamada "Computer Use" que permite al LLM interactuar con la computadora como una persona. Con este agente experimental una persona hablando con el LLM pidió una pizza por internet, el costo por interactuar con el LLM fue de 25 dólares.	Si hay una tarea que por su frecuencia y el tiempo que lleva conviene ser automatizada, en lugar de utilizar un agente experimental con la última funcionalidad igual es más adecuado pedir a un desarrollador crear un agente a medida para esa tarea concreta que minimice el uso de la IA a extraer la información y que luego la parte de introducir esa información en una página web se haga sin la IA
IA locking	Este término lo he derivado a partir de otro llamado "vendor locking" donde es común acabar siendo dependiente de productos de ciertos proveedores, posteriormente por motivos de costo o nuevas funcionalidades resulta muy complicado cambiar de proveedor.	Antes de decantarse por un proveedor concreto, es recomendable comparar la funcionalidad que proporcionan las APIs de otros proveedores y ceñirse a funcionalidades que existan en varios proveedores y adaptar los desarrollos para que en un futuro cambiar de proveedor no sea complicado.
Privacidad	La IA Generativa puede ser muy útil para extraer información o resumir el contenido de documentos que se le proveen. Hay veces que por inconsciencia o bien por necesidad de obtener un resultado rápidamente nos olvidamos que los documentos contienen información confidencial y que al proporcionarlos a la IA estamos haciéndolos públicos.	Los proveedores suelen tener diversos modelos de LLM con mayor o menor grado de privacidad, hay casos en los que se puede indicar que no se usen los datos proporcionados para mejorar sus modelos, en caso de que se retenga la información saber por cuánto tiempo se mantiene y también conocer el coste que conlleva tener más privacidad. En casos que se requiera mayor privacidad se pueden tener modelos LLM de uso exclusivo en datacenters de proveedores o si se quiere tener mayor certeza en servidores propios o ejecutados en local. Para IA en local, no es necesario ni cambiar de ordenador, ya existen aparatos de IA parecidos a un disco duro que se conectan por USB convirtiéndolo en una máquina con una IA potente

La conclusión es que no hay que dejarse cegar por los resultados que ofrecen los nuevos avances y usarlos sin pensar nada más.

Desde su aparición la Inteligencia Artificial Generativa nos ha sorprendido, cada pocos meses ha ido mejorando, por ejemplo las imágenes que genera hoy comparadas con las de hace un año han dado un gran salto de calidad y realismo. Sin embargo se está llegando a los límites de los modelos actuales, por más datos con los que se entrenen los resultados no tienen una mejora tan significativa. Hasta que no se descubran otros modelos más potentes que los LLM no se dará otro salto adelante, mientras se aplicarán a los modelos actuales técnicas como *Chain of Thought* (CoT), *Retrieval Augmented Generation* (RAG) o *Mixture of Experts* (MoE) para intentar mejorar la calidad de los resultados a costa de necesitar mayor capacidad de cálculo y datos.

Mientras al usuario de a pie nos seducirán con la multimodalidad y sobre todo con los agentes que nos harán la vida más fácil automatizando tareas tediosas para nosotros. Vienen años interesantes, pero tengamos dos dedos de frente y no dejemos todo en manos de la IA, al final nos va a pasar como a los que si no les funciona el navegador en el coche se pierden; la IA actual ni es mágica ni es infalible, es mejor usarla como una herramienta de productividad que confiar a ciegas en que hará nuestro trabajo de razonar.

Para sacar el mejor partido hay que usar la Inteligencia Artificial con inteligencia, siendo consciente de las implicaciones de su uso y si no se está seguro de ellas asesorándose con consultores o expertos en el tema. 📄

APROXIMACIÓN CRÍTICA A LA IA GENERATIVA Y SUS SUCEDÁNEOS

Adolfo Plasencia

1. INTRODUCCIÓN

La misma expresión actual de "inteligencia artificial" no significa ya lo mismo que se entendía al nombrarla cuando en 1959 MIT se fundó el Laboratorio de Inteligencia Artificial de MIT con el nombre de Proyecto de Inteligencia Artificial (The AI Lab). Aquel laboratorio fue pionero en el desarrollo de nuevos métodos de cirugía guiada por imágenes y en el acceso a Internet basado en el lenguaje natural; allí se inventaron varias generaciones de micropantallas; y se hicieron realidad las interfaces hápticas; y se desarrollaron robots bacterianos y robots basados en el comportamiento que se utilizan para la exploración planetaria, el reconocimiento militar y en dispositivos de consumo. Ni tampoco significa lo mismo que aquella AI cuya investigación practicaban a principios de los años setenta, en el Laboratorio de Inteligencia Artificial del MIT, Marvin Minsky y Seymour Papert cuando empezaron a desarrollar lo que se conoció como *La Teoría de la Sociedad de la Mente*. En esencia, esa teoría de 1970, intenta explicar cómo lo que llamamos inteligencia puede ser producto de la interacción de partes no inteligentes. Minsky afirma que la mayor fuente de ideas sobre la teoría provino de su trabajo práctico y físico al intentar crear una máquina que utiliza un brazo robótico, una cámara de vídeo y un ordenador para construirlo con bloques infantiles. En 1986, Minsky publicó *La sociedad de la mente*, un libro exhaustivo sobre la teoría que, a diferencia de la mayoría de sus trabajos publicados anteriormente, estaba escrito para el público en general y describir estas ideas pioneras a la gente corriente.

2. LA ETIQUETA SEMÁNTICA 'INTELIGENCIA ARTIFICIAL' (IA)

El uso actual de la etiqueta 'inteligencia artificial' me parece en la mayoría de los casos una corrupción semántica, ya que no es producto de la continuidad del avance a partir de aquellos intentos propios de la curiosidad científica

pura. Dicha etiqueta y su actual significado es producto de estrategias del mundo especulativo financiero ligado a las grandes empresas *big tech* impulsadas por un ánimo de lucro radical de tiempo real, para el que las etiquetas, *machine learning*, o *estadística bayesiana predictiva* –que de eso se trata–, no eran lo suficientemente ‘sexy’ para atraer el dinero financiero ‘grande’ al campo informático, – y, además, con la coartada quizá loable, de acabar con el ‘invierno de la IA’ de principios de siglo–. Así que decidieron relanzar este subcampo informático con esa etiqueta de la informática desplegando una brutal ola de propaganda global sobre los grandes modelos de *software* estadístico predictivo de tipo ‘autocompletador’ pre-entrenados, hiper-hinchados de datos, y anabolizados con una enorme cantidad de computación, que proveerían los *data center* de las *big tech*, hasta tal punto que incluso han obligado por su perentoria necesidad, a considerar como ‘verde’ la energía nuclear. Pero, esto no es fruto de ir en pos del descubrimiento y con la guía por la curiosidad científica, sino especulando, y persiguiendo posiciones a corto como se llama en la jerga, cosa de la que luego mostraré un ejemplo muy ilustrativo en con relación a la IA de Goldman Sachs (GS).

Esas *big tech*, que son quienes nos han metido en pleno siglo XXI en un nuevo medioevo, en una nueva época feudal, como desde hace años asegura¹ el matemático y filósofo Javier Echeverría, quien ya lo auguró en su libro *Los señores del Aire*², que fue Premio Nacional de Ensayo en el año 2000, y al que el tiempo está dando la razón sobradamente.

Linus Torvalds, unos de los líderes pioneros del pensamiento sobre informática, creador del primer kernel Linux, está convencido que la mayor parte del *marketing* difundido por la industria sobre la IA es pura palabrería sin sustancia real, y que pueden pasar muchos años antes de que se demuestre esa tecnología. Durante la Cumbre de Código Abierto de Viena de septiembre

1 Javier Echeverría: «Vivimos en un nuevo feudalismo, pero no de la tierra, sino del aire», Pablo Batalla, El Cuaderno, diciembre, 2017, <https://elcuadernodigital.com/2017/12/18/javier-echeverria-vivimos-en-un-nuevo-feudalismo-pero-no-de-la-tierra-sino-del-aire/>

2 Los señores del aire: Telépolis y el tercer entorno /Javier Echevarría Ezponda. Barcelona, Destino, 1999, <https://digital.csic.es/handle/10261/75365>

de 2024, se le preguntó sobre la IA Generativa, y Linus Torvalds: declaró³ «el 90% del marketing de la inteligencia artificial es publicidad engañosa; creo que la inteligencia artificial es muy interesante y que va a cambiar el mundo, pero al mismo tiempo detesto tanto el ciclo del *hype* [subidón mediático] propagandístico que no quiero entrar en él, así que mi enfoque sobre la inteligencia artificial en este momento es básicamente ignorarla», y añadió: «Creo que toda la industria tecnológica en torno a la IA se encuentra en una posición muy mala; es un 90% marketing y un 10% realidad, y creo que en cinco años las cosas cambiarán y en ese momento veremos de verdad qué parte de la IA se está utilizando para cargas de trabajo reales». Como buen conocedor del sector, explica: «La industria informática es conocida por sus bravatas de *marketing*, que se centran en una tecnología incipiente para prometer más de la cuenta y ofrecer menos de lo esperado». Esto, según el desarrollador del núcleo Linux, es un problema: «Antes de la IA, hace un par de años, de lo único que se hablaba era de criptomonedas. No me gustan esos ciclos *hype*, de bombo y platillo». Torvalds señala que tal es la dimensión global de confusión tecnológica al respecto.

Vayamos por partes. Que esta revolución de la IA Generativa va a tener un impacto muy grande ya no se puede negar. Me lo señalaba en noviembre de 2018, José Hernández-Orallo es catedrático del DSIC, e investigador en el Instituto VRAIN (Valencian Research Institute for Artificial Intelligence) de la UPV. Tuve entonces con este investigador una larga conversación en el Clare Hall College, de la Universidad Cambridge, en la que también es *senior research fellow* del Leverhulme Centre for the Future of Intelligence, además de investigador asociado en el Center for the Study of Existential Risk. José, que también es doctor en lógica y filosofía, –ahora muy relacionadas con la IA–, me dijo expresamente ya hace seis años, que «La inteligencia artificial podría ser la tecnología más transformadora de este siglo».⁴ Y, Hernández-Orallo

3 Linus Torvalds: 90% of AI marketing is hype, Paul Kunert, The Register, 29 Oct 2024, https://www.theregister.com/2024/10/29/linus_torvalds_ai_hype/

4 “La inteligencia artificial podría ser la tecnología más transformadora de este siglo”, 2 noviembre de 2018: https://valenciacity.es/investigacion_y_tecnologia/la-inteligencia-artificial-podria-ser-la-tecnologia-mas-transformadora-de-este-siglo/

sabe de qué habla. Y fue elegido junto a otro miembro de su Instituto VRAIN, –los dos únicos científicos españoles, según Financial Times (FT)⁵–, para formar parte del 'equipo rojo' de científicos de la IA que fueron contratados por la empresa Open AI para «romper el ChatGPT», –en realidad el *Transformer*⁶ GPT-4–. Es decir, según el editor de AI de FT Madhumita Murgia, para «poner a prueba y evaluar posibles riesgos del uso del GPT-4, su entonces nuevo y potente modelo lingüístico».

Esto fue publicado por FT solo un año antes de que Sam Altman el CEO de Open AI, la empresa respaldada por Microsoft, declarara a los cuatro vientos, literalmente, que «The Age of Giant AI Models Is already Over»⁷, es decir que «La era de los modelos gigantes de inteligencia artificial ha terminado». Luego entraré de nuevo en ello. Esa disfunción y contradicción entre hechos, líneas de tiempo y declaraciones, –casi al borde de la disonancia cognitiva–, son típicas y frecuentes, tanto de Altman como de otros famosos evangelistas del *hype* de la IA, ámbito en el que se han hecho ya legendarias sus continuas contradicciones, por su notoria repetición e insistencia. Como si la coherencia no fuera ni siquiera algo a considerar, o como si la audiencia expectante fuera estúpida por naturaleza, y todo le diera igual.

De la repercusión global del subidón mediático (*hype*) de la IA, no hay ninguna duda. Este caso se ha convertido finalmente en la mayor operación propagandística global de la historia, hasta el momento. Y esa repercusión se ha convertido además en una moda digital que parece de obligado cumplimiento. Ello ha provocado, no un río sino un océano de tinta impresa y digital publicada sobre la IA, que ya es imposible leer para cualquier persona; hoy solo sería posible para un *software*, "leer" los millones de textos publicados (por personas o máquinas) sobre la IA en todo el mundo, publicados

5 El equipo rojo de OpenAI: los expertos contratados para "romper" Chat GPT, 14 abril 2023: <https://on.ft.com/3MOcVEt>

6 *Transformer*/Transformador, Wikipedia, [https://es.wikipedia.org/wiki/Transformador_\(modelo_de_aprendizaje_autom%C3%A1tico\)](https://es.wikipedia.org/wiki/Transformador_(modelo_de_aprendizaje_autom%C3%A1tico))

7 OpenAI's CEO Says the Age of Giant AI Models Is Already Over, Wired, 17 abril 2024, <https://www.wired.com/story/openai-ceo-sam-altman-the-age-of-giant-ai-models-is-already-over/>

en internet en los últimos dos años, desde que se lanzó el Chat GPT el 30 de noviembre de 2022, millones de los cuales ya han sido generados por la propia IA Generativa, –lo cual es en sí mismo un meta-proceso–. Tan meta-proceso como lo es el contenido de aquel texto de 32 páginas que recibí y conservo, emitido por el propio Chat GPT en el que esta máquina ‘diagnóstica’ y ‘aconseja’, exactamente, cómo se debía regular la IA Generativa. Que la propia IA Generativa través del Chat GPT, emita un documento de regulación que describe cómo debemos los humanos regularla a ella misma, puede ser interesante siempre que no se le ocurra a ningún descerebrado/a obedecer esa sugerencia. Sobre todo, teniendo en cuenta que, por muy convincente que aparente ser lo que emite, es una máquina que no ‘comprende’ en absoluto el significado de ninguna de palabra de las que emite, ni la diferencia entre verdad y mentira.

Dada la explosión de notoriedad, ya hay en circulación en la red millones de textos sobre IA Generativa ligados o relacionados, con todo tipo de disciplinas porque, como se considera una ‘revolución horizontal’ se dice que la IA va a afectar ‘a todo’. Así que cualquier excusa es buena para publicar sobre el tema de moda y contribuir al *hype*. Una profecía que no tardará en auto-cumplirse, según los llamados genéricamente ‘evangelistas del nuevo Evangelio de la IA’ una expresión que parafrasea al artículo⁸ en que Hans Magnus Enzensberger anticipaba todo esto, ya en el año 2.000, en su texto *El evangelio digital*.

No soy un especialista en IA, pero sí en comunicación tecnológica y científica y en su impacto, en las que tengo muchos años de experiencia. Aún así, no me atrevo a contradecir a Hernández-Orallo sobre si la IA podría ser la tecnología *más transformadora* de este siglo. Suele tener razón.

Por otra parte, lo que aprendí de la IA en el MIT, entre otras circunstancias, sobre todo fue escuchando a Rodney Brooks y Marvin Minsky. En el caso del primero, hablando personalmente con él en su laboratorio. Así que, no me importa confesar que mi forma de ver la IA está bastante influenciada por la

8 El evangelio digital, Hans Magnus Enzensberger, Claves de Razón Práctica N° 104, 2000, 67 <https://dialnet.unirioja.es/servlet/articulo?codigo=151451>

visión del pionero Rodney Brooks, que fue, desde 2004 hasta 2007, director del MIT Computer Science and Artificial Intelligence Laboratory y, además, de fundador de empresas como iRobot, Rethink Robotics y, en 2019, fundó también Robust.AI. Estoy de acuerdo con sus recientes afirmaciones de que «la gente está sobrevalorando enormemente la *IA Generativa*»⁹.

Según la *Ley de Roy Amara*, publicada¹⁰ en 2001 en relación a la evolución de las tecnologías, «Tendemos a sobreestimar el efecto de una nueva tecnología a corto plazo y subestimarla a largo plazo». Esta ley, que ya tiene 23 años, parece describir gran parte de lo que está pasando ahora con la *IA Generativa*.

Revisando cronológicamente dicha evolución reciente, son evidentes las sobreestimaciones sobre los futuros resultados de las tecnologías asociadas a la *IA* que han sucedido sucesivamente. El advenimiento de la *Inteligencia Artificial General* (AGI), ya ha sido fuertemente anunciada, una y otra vez, en sucesivos pasos de avance dados por las tecnologías de *IA Generativa* y sus sucedáneos. Pero, por ahora, no hay tal.

Pondré algunos ejemplos de lo que señala Brooks. McCulloch y Pitts ya hablaron en 1943, por primera vez, sobre la posibilidad de crear *redes neuronales artificiales* como si fueran ordenadores. Frank Rosenblatt, psicólogo estadounidense al que se considera uno de los padres del *Deep Learning*, creó en 1958 el primer algoritmo que presentaba una red neuronal simple que llamó *Perceptrón* (un clasificador binario o discriminador lineal, que genera una predicción basándose en un algoritmo combinado con el 'peso' de las entradas). Para hacerse una idea de hasta qué punto él mismo sobreestimaba ya entonces las expectativas, el propio Rosenblatt escribió en el *New York Times* en 1958, sin cortarse sobre su invento: «este (Perceptrón) –afirmó–, es el embrión de un ordenador electrónico que espera poder caminar, hablar, ver, escribir, reproducirse y tener consciencia de su existencia”.

9 MIT robotics pioneer Rodney Brooks, thinks people are vastly overestimating generative AI, Ron Miller, TechCrunch, June 29, 2024, <https://techcrunch.com/2024/06/29/mit-robotics-pioneer-rodney-brooks-thinks-people-are-vastly-overestimating-generative-ai/>

10 *Amara's Law*, SMART Letter #63, November 26, 2001, <https://isen.com/archives/011126.html>

Casi nadie se salva de la *Amara's Law*. Ni siquiera retroactivamente, Marvin Minsky, que también hizo sus propias sobrestimaciones en 1953 sobre sucesivas tecnologías: con la *lógica de primer orden*; la *lógica difusa* (*fuzzy logic*); el *generador de planes automatizado* STRIPS (Stanford Research Institute Problem Solver); o con las *redes neurales con retro-propagación*. Igual sucedió con los primeros proyectos de coche auto-conducidos propuestos por Dickmanns en 1987; y con las primeras plataformas SOARS, (Security Orchestration and Automation Response). De nuevo, las expectativas de hincharon con los inicios de aprendizaje por refuerzo, el área del *aprendizaje automático* inspirada, en la psicología conductista de Skinner, cuya ocupación es determinar qué acciones debe escoger un agente de *software* en un entorno dado con el fin de maximizar alguna noción de 'recompensa' o premio acumulado. Luego, llegó la *inferencia bayesiana*, o *estadística* en la que las evidencias u observaciones se emplean para actualizar o inferir la probabilidad de que una hipótesis pueda ser cierta. Y también, en ella se exageraron sus posibles logros.

El siguiente salto, fue el del *aprendizaje profundo*, que va más allá de los algoritmos tradicionales de aprendizaje automático. Fue en 2006, cuando Geoffrey Hinton (que acaba de ganar, curiosamente, el Premio Nobel de Física en 2024, siendo él científico computacional, especialista en cognición y en psicología cognitiva). Hinton presentó entonces, hace 18 años, el concepto de *deep learning*, o *aprendizaje profundo*, para explicar unos nuevos algoritmos capaces de 'ver' y 'distinguir' objetos y textos en diferentes imágenes y vídeo. Estos algoritmos tienen una 'capacidad de aprendizaje' independientemente de la cantidad de datos que 'adquieran' y por ello, pueden mejorar su rendimiento al poder acceder a un mayor número de datos, o lo que es lo mismo, hacer que la máquina tenga más "experiencia" –entre comillas– Una vez que las máquinas han adquirido suficiente 'experiencia' [de nuevo, sic] mediante el *aprendizaje profundo*, estos algoritmos pueden ponerse a 'trabajar' [sic, de nuevo] en tareas específicas, en temas como conducir un coche, detectar malas hierbas en un campo de cultivo, detectar enfermedades, inspeccionar maquinaria para identificar errores, etc. Esta era una teoría que generó altas expectativas en la primera década del siglo XXI que, en su mayoría no llegaron a cumplirse.

Después, llegaron las tecnologías de *Watson* en 2011, que le hicieron ganar en el concurso *Jeopardy* (contestando preguntas que le obligaban a «discurrir como las personas», según la descripción); luego, en 2014, DeepMind, –fundada por Demis Hassabis en Londres–, desarrollo el software *AlphaGo* para jugar al juego de mesa *Go* y que en octubre de 2015 se convirtió en la primera máquina de *Go* en ganar a un campeón mundial y jugador profesional del juego, sin emplear piedras de *handicap*.

También en 2014, se produjeron avances como el *autocodificador variacional* y la *red generativa adversativa* que produjeron las primeras *redes neuronales profundas* prácticas capaces de ‘aprender’ *modelos generativos*, en lugar de *discriminativos*, integrados por datos complejos como son las imágenes. Estos modelos generativos profundos fueron los primeros en producir *imágenes generadas* completas. Es decir, no fotografiadas ni filmadas, sino *generadas*, es decir, construidas directamente por algoritmos.

En 2016, se crearon con algoritmos de regresión simbólica, distintas de las *redes neuronales profundas*, a lo que se decidió llamar *scientists machines*, que tuvieron una sorprendente aplicación, las de la ‘física GoPro’, en la que una cámara virtual puede apuntar a un evento y sus algoritmos puede identificar la ecuación física subyacente. Y como señalaba un artículo¹¹ de Quanta magazine, tiempo después, estas poderosas ‘*máquinas científicas*’ [es mi traducción libre], pueden ‘destilar’[sic] las leyes de la física a partir de datos en bruto. En 2017, la red *Transformador*, ya iba equipada con un *tokenizador* de pares de bytes que segmenta la entrada de texto o imágenes en *tokens*. Esto permitió avances en los modelos generativos, lo que llevó al primer *Transformer Generativo Pre-entrenado* (GPT) en 2018. A esto le siguió en 2019 GPT-2, que demostró la capacidad de ‘generalizar’ [sic] sin supervisión a muchas tareas diferentes como modelo fundacional. Y de ahí a los grandes modelos de lenguaje LLM (*Large Language Models*) actuales.

11 Powerful ‘Machine Scientists’ Distill the Laws of Physics From Raw Data, Charlie Wood, Quanta Magazine, May 10, 2022, <https://www.quantamagazine.org/machine-scientists-distill-the-laws-of-physics-from-raw-data-20220510/>

Pues bien, en todas estas sucesivas tecnologías de la IA las expectativas se fueron exagerando al límite y más allá. Un ejemplo muy reciente de esa exageración fue en marzo de 2023, cuando se lanzó GPT-4, de que un equipo de Microsoft Research argumentó que «podría verse razonablemente como una versión temprana (pero aún incompleta) de un sistema de *inteligencia artificial fuerte* (IAF). [de nuevo, sic]. Parece que nada mas lejos de la realidad. El *marketing push* siempre va, empujado, por delante de la realidad científica, sobre todo cuando se trata de ciencia mercenaria, que también la hay, y con mucha financiación oportunista detrás. En resumen, en cada uno de estos pasos, sus expectativas exageradas, anuncian, una y otra vez, la inminencia de una inteligencia de nivel humano en estas máquinas de *software*, dando de nuevo la razón a Roy Amara y su 'ley'.

Pero es obvio que, cuando tienes en la nuca el insistente aliento de los grandes accionistas en lugar de la curiosidad científica, es mucho más difícil tener paciencia, e incluso ser realista y sincero. Brooks advierte que hay que aceptar que siempre va a haber casos atípicos difíciles de resolver cuando se trata de IA. Casos que podrían tardar muchas décadas en resolverse y, añade que existe la creencia errónea, sobre todo gracias a la ley de Moore, de que siempre habrá un crecimiento exponencial en lo que respecta a la tecnología: ... «como la idea de que, si Chat GPT 4 es así de bueno, ¡imagínate cómo serán Chat GPT 5, 6 y 7!». Como si la innovación y el descubrimiento fueran automatismos de velocidad constante, pero no lo son. Él ve un fallo en esa lógica, ya que la tecnología no siempre crece exponencialmente, y por tanto la *Ley de Moore* no es el mejor ejemplo para comparar.

3. POR QUÉ LA IA NO PUEDE RESOLVER TODO. EL SOLUCIONISMO DE LA IA

Quien sí se atreve a polemizar sobre la IA es Vyacheslav Polonski, investigador de la Universidad de Oxford que el mismo año, –justo seis meses antes de mi conversación en Cambridge con Hernández-Orallo–, publicó un artículo que titulaba muy ilustrativamente: «Por qué la AI no puede resolver todo»¹²,

12 Why AI can't solve everything, The Conversation, 25 mayo 2018, <https://theconversation.com/why-ai-cant-solve-everything-97022>

que arrancaba diciendo, ya por entonces, –no lo olvide el lector, en 2018–, en el que afirmaba crudamente:

La histeria sobre el futuro de la inteligencia artificial (IA) está en todas partes. No parecen faltar noticias sensacionalistas sobre cómo la AI podría curar enfermedades, acelerar y mejorar la innovación y creatividad humanas. Con sólo mirar los titulares de los medios de comunicación, se podría pensar que ya estamos viviendo en un futuro en el que la IA se ha infiltrado en todos los aspectos de la sociedad.

Si bien es innegable que la IA ha abierto una gran cantidad de oportunidades prometedoras, también ha llevado al surgimiento de una mentalidad que puede describirse mejor como 'solucionismo de la IA'. Esta es la filosofía de que, supuestamente, con suficientes datos, los algoritmos de aprendizaje automático pueden resolver todos los problemas de la humanidad.

Y auguraba, además, en ese texto, ya hace seis años:

En tan sólo unos años, el 'solucionismo de la IA' ha pasado de las bocas de los evangelistas de la tecnología en Silicon Valley a las mentes de los funcionarios gubernamentales y políticos de todo el mundo. El péndulo ha pasado de la noción distópica de que la IA destruirá a la humanidad a la creencia utópica de que nuestro salvador algorítmico está aquí....

Ahora estamos viendo que los gobiernos se comprometen a apoyar las iniciativas nacionales de IA y compiten en una 'carrera armamentista' tecnológica y retórica para dominar el floreciente sector del aprendizaje automático. Por ejemplo, el gobierno del Reino Unido se ha comprometido a invertir 300 millones de libras esterlinas en la investigación de la IA para posicionarse como líder en este campo. También, enamorado del potencial transformador de la IA, el presidente francés Emmanuel Macron se comprometió a convertir a Francia en un centro mundial de IA. Mientras tanto, el gobierno chino está aumentando su capacidad de inteligencia artificial con un plan

nacional para crear una industria china de inteligencia artificial por valor de 150.000 millones de dólares para 2030. El 'solucionismo de la IA' está en ascenso y está aquí para quedarse...

Pero hay un gran problema con esta idea. En lugar de apoyar el progreso de la IA, en realidad pone en peligro el valor de la inteligencia de las máquinas al hacer caso omiso de importantes principios de seguridad de la IA y establecer expectativas poco realistas sobre lo que la IA puede hacer realmente por la humanidad.

El profesor Polonski, consciente de que cuando publicó su artículo ya había un enorme contagio del *pensamiento mágico* de la IA, que, –como dijo él mismo–, ya estaba aquí para quedarse. Advertía también que el *aprendizaje máquina*, –rebautizado ahora por extrapolación casi en la mayoría de casos, directamente como "IA" –, no es mágico, y que...

Si queremos cosechar los beneficios y minimizar los daños potenciales de la IA, debemos empezar a pensar en cómo el aprendizaje automático puede aplicarse de manera significativa a áreas específicas del gobierno, las empresas y la sociedad. Esto significa que necesitamos tener una discusión sobre la ética de la IA y sobre la desconfianza que crecientemente mucha gente tiene hacia el aprendizaje automático...

Y lo que es más importante, debemos ser conscientes de las limitaciones de la IA y de los aspectos en los que los seres humanos aún deben tomar la iniciativa. En lugar de pintar una imagen poco realista del poder de la IA, es importante dar un paso atrás y separar las capacidades tecnológicas reales de la IA, de la magia.

Y, culminaba su reflexión afirmando que:

Usar la IA simplemente por el bien de la IA, no siempre puede ser productivo o útil. No todos los problemas se resuelven mejor aplicando "inteligencia de máquina" a ellos. Esta es la lección crucial para todos los que quieren impulsar las inversiones en los programas nacionales de IA: todas las soluciones tienen un coste y no todo lo que se puede automatizar debería tenerlo.

4. MACHINE LEARNING; REPLICAR Y AUTOMATIZAR EL APRENDIZAJE HUMANO

Desde el punto de vista de la ciencia de la computación, hace ya décadas, la era digital ha acelerado los intentos de conseguir un *software* o máquina capaces de realizar tareas que, sobre todo, sean repetitivas. Eso ha dado lugar a la llamada *automatización*, de forma que actuaciones en procesos de producción que eran realizados por humanos van siendo sustituidos cada vez más por sistemas robóticos físicos o sistemas de *software*, de forma que este proceso ha eliminado muchos puestos de trabajo e incluso profesiones de trabajadores humanos. Durante décadas se ha hablado de que iban a sustituir a trabajadores humanos en funciones 'repetitivas' y no lo harían en trabajos con funciones 'creativas', pero esa línea roja ya está siendo superada¹³ de forma creciente y nihilista por el afán de reducir costes al precio que sea, cuanto antes y sin considerar las consecuencias.

La vertiginosa evolución de la informática ha derribado esa visión, poniendo en uso a gran escala los desarrollos o modelos de *sistemas algoritmos de estadística computacional masiva*, capaz de desplegar el *aprendizaje automático*, o *aprendizaje automatizado*, también llamado *aprendizaje de máquinas* (del inglés, *Machine Learning. ML*) que es un subcampo de las ciencias de la computación y de la inteligencia artificial, cuyo objetivo es desarrollar técnicas que permitan que el *software* 'aprenda' aunque este término se usa la mayoría de las veces recurriendo al antropomorfismo. Pero si antropomizáramos ya sin complejos podríamos decir que en el 'aprendizaje de máquina' un *software* 'observa' y lee datos, luego 'construye' un modelo basado en esos datos y utiliza ese modelo a la vez como 'hipótesis digital acerca del mundo' cercano. Mediante este sistema una 'pieza' de *software* que puede resolver problemas cada vez más distintos, siendo optimistas.

En cuanto a resolver tareas concretas con éxito, el *aprendizaje automático* ya tiene una amplia gama de aplicaciones, incluyendo motores de búsqueda, diagnósticos médicos, detección de fraude en el uso de tarjetas de crédito,

13 Is AI replacing journalists? One broadcaster replaced its announcers with new ones powered by artificial intelligence, The Associated Press, 23 octubre, 2024, <https://apnews.com/article/poland-media-artificial-intelligence-radio-cbae50063b8f125c99a4a0f6d3ee82d8>

análisis de mercado para los diferentes sectores de actividad, clasificación de secuencias de ADN, valoración de riesgos, o reconocimiento del habla y del lenguaje escrito, o en juegos y robótica. Sin embargo, en todas estas aplicaciones la solución llega mediante la búsqueda de patrones y relaciones de los sistemas algoritmos, por ejemplo, de *estadística masiva de tipo bayesiano*, o de *'matching'*¹⁴ basada en una enorme magnitud de computación y en el poder de los datos, en el contexto del 'Big Data'.

Sin embargo, estos sistemas de *software*, hoy de uso masivo, no son verdadera inteligencia artificial (IA), sino solo *Machine Learning (ML)*, según el experto Francisco J. Martín¹⁵, co-fundador y CEO de la exitosa empresa BigML, que afirma que es así en más del 90% de los casos, aunque en todos ellos se les suele llamar *aplicaciones de inteligencia artificial*.

En la última década el campo de la inteligencia artificial está explotando a la vez en muchos frentes de áreas especializadas concretas. Y cada vez más, estos sistemas de *software* se están ocupando de funciones más y más creativas. La *Inteligencia Artificial (IA)* abarca en la actualidad una gran variedad de subcampos, que van desde áreas que buscan el propósito general (como los que está intentando la empresa DeepMind¹⁶, que lidera Demis Hassabis, Premio Nobel de Química 2024, y hoy en Google), a la gran algorítmica de *Machine Learning* especializado en *Big Data*, que aplican en sus actividades comerciales y publicitarias casi todas las grandes plataformas y empresas tecnológicas.

14 En informática estadística, la técnica de 'matching' se lleva a cabo con funciones de emparejamiento es una relación matemática que describe o señala la formación de nuevas relaciones (también llamadas 'emparejamientos') de entidades o conjuntos de datos no emparejados, de los tipos adecuados.

15 Francisco J. Martín: "¿Abrir el grifo en una empresa y que salga IA? Ya estamos cerca, Adolfo Plasencia, Valencia Plaza, 18 de octubre de 2021, <https://castellonplaza.com/francisco-j-martin-abrir-el-grifo-en-una-empresa-y-que-salga-ia-ya-estamos-cerca>

16 DeepMind encabeza la carrera por una inteligencia artificial general, Adolfo Plasencia, El Español, 2 octubre 2028, https://www.elespanol.com/invertia/disruptores-innovadores/innovadores/investigacion/20181002/deepmind-encabeza-carrera-inteligencia-artificial-general/342467424_o.html

Asegura el discurso de marketing actual de la IA que Esta *estadística computacional predictiva* a gran escala, que es ML puro y duro, ya muy especializada, se empieza a usar en campos intelectuales creativos complejos, y generalizados, como el aprendizaje y la percepción, o más específicas como el juego de ajedrez o del *Go*; en la búsqueda demostración de teoremas matemáticos complejos mediante *algoritmos de regresión simbólica*; la realización de espectaculares piezas audiovisuales llamadas *falsificaciones profundas*¹⁷. Todo ello eran tareas netamente intelectuales realizadas hasta hace muy poco sólo por seres humanos, como ocurría con la escritura de literatura, o poesía; y el diagnóstico de enfermedades. Pero hoy están siendo 'asumidos', -de nuevo, antropoformizando-, crecientemente por la *Inteligencia Artificial Generativa* (IAG) que ya 'sintetiza' y 'automatiza' esas tareas esencial y estrictamente intelectuales. Así que, por lo tanto, la IA 'avanzada' -es un calificativo que ya se usa también en la nueva *jerga* IA, -aunque no sabemos de qué grado de avance se habla-, es ya potencialmente relevante para cooperar o 'asumir cualquier ámbito' [sic] de la actividad intelectual humana. Esto lo dicen con seguridad los defensores radicales de la nueva IA.

5. ¿PUEDE LA IA 'ESCRIBIR' POESÍA AUTÉNTICA?

Con respecto a la poesía, la literatura y su relación con la IA, el psicólogo cognitivo y poeta Keith Holyoak, autor del libro *The Spider's Thread Metaphor in Mind, Brain, and Poetry*¹⁸ (*El hilo de la araña. La metáfora en la mente, el cerebro y la poesía*), se muestra predispuesto a aconsejar la IA Generativa planteada como un conjunto de recursos de ayudas 'técnicas' para escritores y poetas. Y dando un paso más preguntaba en el título de un artículo suyo de MIT Press Reader: «¿Puede la IA escribir poesía auténtica?»¹⁹. Pero esa pregunta plantea otras preguntas muy subjetivas. ¿Lo que emita o 'escupa' una

17 Falsedades digitales profundas, Adolfo Plasencia, El Español, 24 enero 2019, https://www.elespanol.com/invertia/disruptores-innovadores/opinion/20190124/falsedades-digitales-profundas/371082891_12.html

18 *The Spider's Thread Metaphor in Mind, Brain, and Poetry*, Keith J. Holyoak, (MIT Press), <https://mitpress.mit.edu/9780262551472/the-spiders-thread/>

19 Can AI Write Authentic Poetry?, Keith J. Holyoak, The MIT Press Reader, <https://thereader.mitpress.mit.edu/can-ai-write-authentic-poetry/>

IA, es decir, un Modelo o *loro estocástico*, ¿podrá ser considerado verdadera poesía? Será, supongo, que como acción de respuesta de un *Modelo LLM*, pero creo que es mucho imaginar que una IA 'quiera' hacer poesía mientras no este dictada de intención, aunque eso se podría simular. Aunque, que yo sepa, las IA generativas actuales carecen de voluntad, solo cumplen objetivos que algún humano les ha pre-programado. Y, si carecen de ella, ¿cual cree el lector que sería la respuesta a la pregunta del titulo del citado artículo? ¿Y quien decide si lo que sale de la *Caja negra* del LLM es 'poesía auténtica o falsa'? ¿Los lectores? ¿Los críticos? ¿El mismo LLM que la escupa?

De todas formas, Holyoak está por la labor de usar la IA Generativa en el campo literario Y experimentar. Y hace afirmaciones atrevidas en su artículo: «En cierto sentido, la poesía puede servir como una especie de canario en la mina de carbón, un indicador temprano de hasta qué punto la IA promete desafiar a los humanos como creadores artísticos»; y como poeta optimista, explica a sus colegas interesados (colegas lo de poetas, creo, pero no necesariamente en lo de psicólogo cognitivo), que ahora «Además de tener acceso a un diccionario de sinónimos y de rimas automatizado, un poeta puede recurrir a un sugeridor de metáforas informatizado». En su artículo, resume los argumentos de su libro. Pero, finalmente en el artículo, –ni en el libro–, en realidad no es capaz de responder a la pregunta de su propio título. A lo más que llega es a hacerse más preguntas, optimistas como: «Considerados como textos, ¿alcanzarán los poemas de IA algún nivel que pueda confundirse con la grandeza humana, o la poesía creada sin experiencia interior presionará inevitablemente en vano contra algún límite inherente» ... «¿Podría la IA crear metáforas realmente nuevas, en lugar de limitarse a variaciones de las que ya hemos inventado los humanos? Esto no lo había visto yo hasta ahora. Creo que es la *Ley de Roy Amara* aplicada a la poesía.

Hice ya, a raíz de citado libro, hace un tiempo mi propia reflexión en 2022, sobre la posición de los creativos literarios en relación a la IA titulada: *Editores, escritores y poetas ante la inteligencia artificial* ²⁰ a raíz de un experimento

²⁰ Editores, escritores y poetas ante la inteligencia artificial, Adolfo Plasencia, Valencia City, 31 diciembre 2022, <https://valenciacity.es/actualidad/editores-escriitores-y-poetas-ante-la-inteligencia-artificial/>

del editor de MIT Press, que le encargó una carta de presentación de un libro 'superventas' imaginario al Chat GPT y le pidieron también unos bocetos de la portada. No lo han vuelto a hacer. Pero, sinceramente, mi reflexión no es tan optimista como la de Holyoak. Por otra parte, la editorial ya no ha publicado ningún experimento con el Chat GPT y de aquella prueba van a hacer dos años.

Después, se ha descubierto posteriormente que ciertos medios que han publicado artículos hechos con IA, pero firmados por personas, que se han considerado un fraude. Por ejemplo, The_Byte, publicó²¹ que la importante web de tecnología CNET ha publicado decenas de artículos generados con *inteligencia artificial*. No se trataba de la automatización de algún párrafo. Algún editor en la empresa, dio luz verde para que se 'escribieran' así en su totalidad. La investigación confirmó que más de 70 artículos firmados bajo el seudónimo CNET Money, todas enfocadas en consejos financieros y, 'presuntamente verificados' por personas del equipo editorial. Fue un escándalo enorme que obligó a poner un aviso con una descripción desplegable en cada artículo que aún se puede ver en la web y que señala: «este artículo ha sido generado mediante tecnología de automatización». También investigué personalmente el uso de la IA en varios medios de EE.UU. que publiqué en un texto titulado: *La invasión algorítmica se apodera del periodismo* ²², en la que describo cómo en los grandes medios de EE.UU. se está usando la IA a nivel industrial para multiplicar la producción de piezas sobre noticias. En ese momento también se descubrió que Amazon vendía y vende libros guías de turismo escritas por autores imaginarios. Llevan fotos en portada de estos autores inexistentes que están generados por *IA Generativa*. Tras un breve periodo de escándalo –eso en el mundo editorial se consideran un fraude–, continúan vendiéndose ya que parece que no lo es para Amazon.

Así que en estos ámbitos editoriales y literarios donde la *inteligencia artificial* desplegada en forma de poderosos sistemas algorítmicos, están dando

21 CNET is quietly publishing entire articles generated by ai, Frank Landymore, 15 enero 2023, <https://futurism.com/the-byte/cnet-publishing-articles-by-ai>

22 La invasión algorítmica se apodera del periodismo, Adolfo Plasencia, 19 agosto de 2023, <https://adolfoplasencia.es/blog/la-invasion-algoritmica-se-apodera-del-periodismo/>

grandes sorpresas y también nuevos sobresaltos. Pero yo creo que, en sus costumbres tecnológicas actuales, la gente en la práctica ya no se hace preguntas grandes, y ni siquiera pequeñas, sobre ello, y compra las citadas guías por internet. Parece que les da igual que sean falsas, o ni lo saben.

Otros prefieren ponerle *prompts* al *Chat* al que han decidido subcontratar sus preguntas, o mejor, sus respuestas y luego hacerlas pasar por suyas sin cargo de conciencia alguno. Ya como costumbre generalizada, la *comodidad nihilista y el conformismo tecnológico*²³ se ha impuesto. Y, bastantes usuarios, para ahorrar tiempo y esfuerzo mental, le subcontratan la expresión de sus conocimientos al *software de IA*, y también hacen simulando su 'autoría' con toda naturalidad. Hay numerosas pruebas en el mundo literario, de que la *IA Generativa* se ha convertido en una herramienta de fraude a nivel industrial. Disimulen. Pero la detección del fraude también se está convirtiendo en otro nuevo negocio.

Por cierto, la prestigiosa *Nature*, que no es de poesía, sino de ciencia, que cambió súbitamente sus reglas de publicación ante los primeros fraudes de autoría mediante *IA Generativa*, después, se ha dado cuenta de que esta tecnología podía ser buena para el negocio, –no se puede ir en contra el 'imparable progreso'–, y ha creado un nuevo epígrafe llamado *Nature Machine Intelligence* con unas reglas específicas denominadas *AI-assisted writing-Nature*. El negocio es el negocio. Y si lo publica *Nature*, es ciencia, faltaría más. No hace falta preguntarle eso al Chat GPT.

6. EL ORÁCULO FINANCIERO POR EXCELENCIA, SE PRONUNCIA Y SE DES-PRONUNCIA SOBRE LA IA (GENERATIVA)

Pero no solo los usuarios de a pie incurren en contradicciones. También los ámbitos especializados que actúan como grandes medios de persuasión. Veamos un ejemplo. Dentro de la vorágine del mundo financiero, de entre los principales impulsores globales como etiqueta semántica de la expresión "AI", está Goldman Sachs (GS) el grupo de banca de inversión y valores más grande

23 Comodidad nihilista y conformismo tecnológico, Valencia City, 20 febrero 2024: <https://valenciacity.es/articulos-de-opinion/comodidad-nihilista-y-conformismo-tecnologico/>

del mundo. Está especializado en elucubrar con éxito, –dadas las informaciones a las que tiene acceso–, ya que debe dar orientaciones públicas a sus lectores, muchos de ellos operadores de bolsa, inversores o economistas. En este caso, la mayoría de los lectores de la enorme audiencia que tiene, piensan que es *de facto* uno de los principales oráculos capitalistas del mundo. Su división que actúa como *oráculo* económico se llama Goldman Sachs Research que el 5 de abril de 2023 publicó un rotundo informe con afirmaciones dignas de titulares de primera página cosa que, como de costumbre, en ellos situaron bastantes medios de inmediato en aquel momento ya que muchos editores, en la prensa económica global, dan por buenas las afirmaciones de este *oráculo*, como si de las de un evangelio financiero se tratara, es decir, casi de obligada creencia. El explosivo titular decía: «La IA generativa podría elevar el PIB mundial un 7%»²⁴, y en su contenido, –que muchos medios suelen repetir sin sus “podría” dándolos por hecho–, afirmaba:

...los avances en inteligencia artificial generativa pueden provocar cambios radicales en la economía mundial. A medida que las herramientas que utilizan los avances en el procesamiento del lenguaje natural se abren camino en las empresas y la sociedad, podrían impulsar un aumento del 7% (o casi 7 “billones” de dólares) en el PIB mundial y elevar el crecimiento de la productividad en 1,5 puntos porcentuales en un período de 10 años.

En dicho informe, los economistas de Goldman Sachs (GS) Joseph Briggs y Devesh Kodnani explican: «A pesar de la gran incertidumbre que rodea al potencial de la IA generativa, su capacidad para generar contenidos indistinguibles de los creados por los humanos, y para romper las barreras de comunicación entre humanos y máquinas, reflejan un gran avance con efectos macroeconómicos potencialmente importantes». Dichos economistas añaden que «Una nueva oleada de sistemas de IA también puede tener

24 Generative AI could raise global GDP by 7%, Goldman Sachs, 5 abril 2024, <https://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent.html>

importantes repercusiones en los mercados de trabajo de todo el mundo. Los cambios en los flujos de trabajo provocados por estos avances podrían exponer a la automatización el equivalente a 300 millones de empleos a tiempo completo». Todo ello, acompañados de gráficos proverbialmente convincentes para reforzar sus afirmaciones y que 'entren por los ojos', casi sin pensar.

Sin citar a Schumpeter y su teoría de la "destrucción creativa", estos economistas le parafrasean y afirman que, «además, los puestos de trabajo desplazados por la automatización se han compensado históricamente con la creación de otros nuevos, y la aparición de nuevas ocupaciones tras las innovaciones tecnológicas representa la gran mayoría del crecimiento del empleo a largo plazo». Según el informe, por ejemplo, «las innovaciones en tecnologías de la información introdujeron nuevas ocupaciones como diseñadores de páginas web, desarrolladores de *software* y profesionales del *marketing* digital. Esta creación de empleo también tuvo efectos secundarios, ya que el aumento de la renta agregada impulsó indirectamente la demanda de trabajadores del sector servicios en industrias como la sanidad, la educación y los servicios de alimentación». A su vez, otro informe complementario de Goldman Sachs Research afirma: «...Al mismo tiempo, se espera que los avances en IA tengan implicaciones de gran alcance para los sectores del *software* empresarial, la sanidad y los servicios financieros.» Y reiterando las afirmaciones corporativas de la empresa, en un tercer informe, del equipo de GS, Kash Rangan, analista sénior de *software* estadounidense en Goldman Sachs Research, explica «La tecnología se está abriendo paso en las aplicaciones empresariales, mejorando la eficiencia cotidiana de los trabajadores del conocimiento, ayudando a los científicos a desarrollar fármacos con mayor rapidez y acelerando el desarrollo de código de *software*, entre otras cosas.»

Y señala, además, «Las empresas de *software* ya están equipando sus carteras de productos con nuevas ofertas de *IA Generativa*. Las empresas de *software como servicio (SaaS)*, por ejemplo, lo están utilizando para abrir oportunidades de venta cruzada de productos y aumentar la retención y expansión de sus clientes, señalan los autores. Estas empresas pueden aprovechar la IA generativa de varias formas para crecer: 1) con nuevos lanzamientos de producción y aplicaciones, 2) cobrando primas por ofertas integradas en la IA, y 3) aumentando los precios con el tiempo a medida que los

productos existentes se complementan con funciones habilitadas para la IA y demuestran su valor a los clientes». Y, finalmente, el economista de GS acaba dando cifras también: «...el mercado total al que puede dirigirse el *software* de *IA Generativa* es de 150.000 millones de dólares, frente a los 685.000 millones de dólares de la industria global del *software*».

Parece incontestable. ¿Verdad? Todo eso lo publicó oficialmente²⁵ Goldman Sachs el 5 de abril de 2023.

Pues bien, atención. La misma empresa GS, el mismo poderoso *oráculo* económico, solo quince meses después, el 12 de julio de 2024, publicó otro demoledor informe, no menos rotundo, cuyo titular se puede resumir en la potente frase: «La IA está sobrevalorada, es cara y poco fiable». Lo dice en un documento de investigación que cuestiona rotundamente la viabilidad económica de la *IA Generativa*, en el que señala que «hay poco que mostrar» en relación a la enorme cantidad de gasto hecho en infraestructura de *IA Generativa*. El texto citado cuestiona y se pregunta «si este gran gasto se amortizará alguna vez en términos de beneficios y rendimientos de la IA». El voluminoso informe, tiene como núcleo principal, un *paper* de investigación titulado: «Gen AI: too much spend, too little benefit?»²⁶, es decir: «AI Generativa: ¿demasiado gasto y pocos beneficios?». La publicación se basa en una serie de entrevistas con economistas e investigadores de Goldman Sachs, y el profesor del MIT Daron Acemoglu, hoy flamante Premio Nobel de Economía 2024, complementada con el análisis de algunos expertos en infraestructuras.

En resumen, el documento cuestiona fuertemente que la *IA Generativa*, –como, por ejemplo, la del tipo de los Modelos como la familia Chat GPT, o Dall-E, LLaMA, Claude, Bard, Watsonx AI, Stable Diffusion, Midjourney, Gemini, etc. y otros similares–, llegue a convertirse en la tecnología transformadora por la que apuestan actualmente Silicon Valley y gran parte del mercado

25 La IA Generativa podría elevar el PIB mundial un 7%, Goldman Sachs Research, 5 de abril de 2023, <https://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent.html>

26 Gen AI: too much spend, too little benefit, <https://www.goldmansachs.com/intelligence/pages/gs-research/gen-ai-too-much-spend-too-little-benefit/report.pdf>

bursátil. Pero, además afirma, –creo que irónicamente y entre líneas, ya que sabemos que ese es el objetivo de GS y no la ciencia–, que los inversores «pueden seguir enriqueciéndose de todos modos» ... Y, una vez esto claro, continua... «a pesar de estas preocupaciones y limitaciones, todavía vemos margen para que el tema de la IA siga su curso, bien porque la IA empiece a cumplir sus promesas, bien porque las burbujas tarden mucho en estallar». Diáfano.

Los investigadores de Goldman Sachs insisten, ahora, –no debe importar, al parecer, a sus lectores del informe anterior que el nuevo contradiga, radicalmente, lo que decía la misma empresa hace solo unos meses–, que... «el optimismo ante la IA está impulsando un gran crecimiento de valores como Nvidia y otras empresas del S&P 500 (las mayores empresas del mercado bursátil)». Para poner la guinda y dejar tranquilo al lector de su informe afirman que «...las ganancias en el precio de las acciones que hemos visto se basan en la suposición de que la *IA Generativa* va a conducir a una mayor productividad (lo que necesariamente significa automatización, despidos, menores costes laborales y mayor eficiencia)». O sea, las anteriores suposiciones eran equivocadas y de novatos sobre el tema, al parecer.

Y, finalmente, parece que para tranquilizarte por si eres alguien del mundo financiero que sigues sus consejos, Goldman Sachs (GS) dice en su nuevo documento de 2024 que «las ganancias bursátiles ya están previstas». Y envía otro mensaje (también fuertemente contradictorio con su informe de pocos meses antes). Aunque parece no importar, al parecer porque los creyentes creen al oráculo a pies juntillas, que aclara: «Aunque el repunte de la productividad que promete la IA podría beneficiar a la renta variable a través de un mayor crecimiento de los beneficios, observamos que las acciones a menudo anticipan un mayor crecimiento de la productividad antes de que se materialice, lo que aumenta el riesgo de pagar en exceso. Y utilizando nuestro nuevo marco de previsión de rendimientos a largo plazo, descubrimos que puede ser necesario un escenario muy favorable de IA para que el S&P 500 ofrezca rendimientos superiores a la media en la próxima década». Estupendo.

En palabras llanas lo que esto supone es que una de las mayores instituciones financieras del mundo está observando que lo que la gente sigue es la inercia del gigantesco *hype*. La potente propaganda financiera ha hecho

su efecto y las empresas y la enorme masa de fans de *la IA Generativa* ya están actuando *de facto* como si esa IA fuera a cambiar el mundo irremediablemente. Y, por eso están actuando masivamente como tales, mientras que la realidad, según la misma GS, es que se trata actualmente una tecnología «profundamente poco fiable», en palabras literales de GS, a fecha 25 de junio de 2024 y que «puede que no cambie mucho de nada en absoluto». Ni más ni menos.

Pero no importa, Goldman Sachs no ha hecho este demoledor informe para los entusiasmados fans del Chat GPT que ya están enganchados a sus maravillas, ya son creyentes de la religión *gepetera* y como tales actúan, ya que además presos de su nueva creencia siguen convenciendo a todo el que pueden que esta tecnología ha cambiado sus vidas para siempre y lo va a hacer igual por las empresas que conocemos en todo el orbe, sean grandes o pequeñas. Así que bienvenidos sean esta inmensa nueva y súbita cohorte y multitud de convencidos de la *IA Generativa* o de sus múltiples sucedáneos, en cuyas tribunas y púlpitos hay sitio para todos y para todas, no importa su número.

La operación especulativa del mundo financiero con la *IA Generativa*, puro *machine learning* para entendernos, ha sido un éxito. Las cotizaciones de las acciones se siguen disparando gracias a todo este *hype* que va a bombo y platillo, a pesar de la realidad, o de que las inversiones ya realizadas, en última instancia, “puede que no cambien casi nada” las cosas. Ahora, a mucha gente, le importa mucho más su emoción y entusiasmo que cualquier otra cosa, por mucho que la situación no se tenga en pie, al decir de GS, ni sea racional. Esto del *irracionalismo* de la IA también va unido a sesgos cognitivos, sobre todo al sesgo de confirmación²⁷. Una vez generada la creencia, ese sesgo, la refuerza más una y otra vez. Como ocurre en las cámaras de eco de las redes sociales.

La odisea de desarrollo del Chat GPT comenzó casi silenciosamente en la empresa. Open AI, creada inicialmente en 2015 como una entidad sin ánimo

27 El poderoso y a veces nefasto sesgo de confirmación, Adolfo Plasencia, 20 octubre 2014, <https://valenciacity.es/articulos-de-opinion/el-poderoso-y-a-veces-nefasto-sesgo-de-confirmacion/>

de lucro –de ahí su nombre, que es lo único que queda de esa idea original–, por Ilya Sutskever, Greg Brockman, Wojciech Zaremba, Andrej Karpathy, John Schulman, Elon Musk y Sam Altman, En aquél 2015 valía 230.000 dólares. Ahora, según Bloomberg²⁸, a 11 de septiembre de 2024, alcanzó un valor económico (bursátil) de 150.000 millones de dólares. Eso es lo que importa. El objetivo del *hype* provocado por los especuladores ha tenido un éxito espectacular en cuanto a su objetivo económico. Pero yo creo que en lo de generar una nueva creencia en la IA, para dejar atrás el invierno de la disciplina, aún lo ha tenido más. Hoy, ya hay pistas de que los mejores cerebros de la empresa Open AI han abandonado el barco, una vez cobrados los dividendos. Por ejemplo, Ilya Sutskever, que era el científico jefe de Open AI, coinventor, junto con Alex Krizhevsky y Geoffrey Hinton (Premio Nobel de Física 2024), de *AlexNet*, una *red neuronal convolucional*.

Como anécdota graciosa, Geoffrey Hinton en la rueda de prensa para anunciar que había ganado el Nobel 2024 declaró que estaba orgulloso de que un ex-alumno suyo, Ilya Sutskever, se atreviese a despedir a Sam Altman antes de irse de Open AI. Hoy, Sutskever ya ha fundado otra empresa de IA llamada Safe Superintelligence. Junto con él otros tres destacados investigadores y grandes cerebros de Open AI también han abandonado la empresa. ¿Importa eso a los creyentes de la IA y del Chat GPT? En absoluto. Probablemente, ni lo sepan. Siguen hablándole y escribiéndoles *prompts* al Chat GPT y a otros *loros estocásticos* como si tal cosa, encantados de subcontratar su corteza cerebral y su pensamiento. Es lo que dice Andrej Karpathy²⁹ el ex-director de inteligencia artificial y *Autopilot Vision* en Tesla, –que después también estuvo en Open AI, a la que ya abandonó–. Dice que disponer de un *exocórtex* para subcontratar tu cerebro es guay. Supongo que para él y para algunos más, lo es, sin duda.

28 OpenAI Fundraising Set to Vault Startup's Valuation to \$150 Billion, Bloomberg, 11 septiembre 2024, <https://www.bloomberg.com/news/articles/2024-09-11/openai-fundraising-set-to-vault-startup-s-value-to-150-billion?>

29 Andrej Karpathy... "we expand our brains into an exocortex on a computing substrate..." <https://twitter.com/tsarnick/status/1831801245549654169>

La batalla por el Relato de la IA parece ya perdida por la ciencia, frente a la propaganda y el *marketing*.

7. CORRUPCIONES SEMÁNTICAS Y TRAMPAS EN LOS DESARROLLOS DE LA IA

La primera trampa de la actual IA es llamarle lo que no es. Recuerdo lo irónico que resultó hace tiempo el acordar con mi amigo, el filósofo Daniel Innerarity, que me aconsejó, que como la magnitud del subidón mediático era tan grande, ...«resulta ya inútil decirle a la gente que eso que llaman IA no es lo que dicen que es». Pero en cualquier caso, eso es un problema que va a continuar porque las necesidades propagandísticas, –en realidad intereses creados artificialmente y de raíz económica–, suelen intentar imponerse, incluso realizando afirmaciones ontológicas falsas que, cuando los hechos van demostrando que no son ciertas, o pierden fuelle publicitario acuden a la polisemia sucesiva, ya que la como parece que la IA “auténtica” definitiva no llega por la vía de los hechos, se multiplican las nuevas denominaciones o etiquetas semánticas, e incluso las palabras clave clicables o esos *hashtags* del internet social que ya procesan masivamente las máquinas.

Así las expresiones van migrando de la de *IA Generativa*, a las de *IA ‘fuerte’* o *‘débil’*, la *IA ‘general’*; la *IA ‘completa’* y, así sucesivamente. Esta confusión en el significado auténtico que no está respaldada por los hechos de la IA, es grave, una auténtica farsa, ya que es de tal magnitud el boom propagandístico, que impone modas *fake* y, por este camino, empiezan a cundir la polisemia interesada y a denominarse con el ‘de IA’ a carreras universitarias, seminarios, ámbitos especializados, múltiples eventos varios, campañas gubernamentales e incluso leyes que, finalmente, no están señalando cosas reales, y como si todos los contenidos hablaran de lo mismo. La realidad a medio plazo no es así, ya que son denominaciones o significados con fecha de caducidad, lo cual conlleva muchos engaños, confusiones y a muchas pérdidas de tiempo, a mucha gente. Yo creo que llamar a algo lo que no es, no es una buena solución.

Aunque la citada primera trampa ha sido una corrupción semántica sobrevenida; dado que el *mainstream* ha adoptado masivamente esa etiqueta semántica de la IA, casi todo el mundo está ya fuertemente evangelizado en

una sociedad como la actual permeada por la propaganda, y por el sesgo de confirmación, creo que no es realmente práctico oponerse radicalmente.

La siguiente trampa es la materia prima que impulsa la estadística predictiva ML, que son los datos y los metadatos de los usuarios de los que, mediante diversos trucos, incluso legales, las plataformas y las *big tech* se apropian sin resistencia mediante *cookies* y *Apps*, verdaderos agujeros negros y sumideros de succión y apropiación de datos y metadatos. Una apropiación, ante la que la gente corriente está indefensa. El instrumento principal para ello son los *Smartphones*, que con su combinación de *hardware* y *software* de *estadística predictiva hiper-segmentada*, constituye hoy una trampa cognitiva casi infalible.³⁰

Esa combinación de la estadística predictiva que fluye continuamente a las pantallas desde las plataformas impulsada por el uso generalizado, casi universal, de los *Smartphones* constituye hoy en día la infraestructura global de manipulación de conductas más grande jamás creada. Y su impacto es enorme, ya que afecta al 62,3% de la población mundial. Según cifras a enero de 2024, del estudio *Digital 2024; Gobar Overview Report*, las plataformas de redes sociales, en gran parte impulsadas por diversas formas de IA, están siendo usadas por 5.040 millones de personas en el mundo. Ni en sus más distópicas predicciones George Orwell pudo imaginar un instrumento de manipulación de esa magnitud. Y mucho menos que dicha manipulación de conducta sería alimentada por los propios manipulados gracias a la continua activación, personalizada y ubicua (en cualquier momento y lugar donde se encuentren), de su circuito cerebral de recompensa instantánea que, además, genera usuarios adictos de diversos grados, en gran parte manipulados y aparentemente felices gracias a su comodidad nihilista y, sobre todo, al conformismo tecnológico³¹ que genera la sensación de que, esta apabullante tecnología ubicua y sus efectos, son de todo punto inevitables. ¿Deberemos

30 Una trampa cognitiva casi infalible, Adolfo Plasencia, Valencia City, 15 de marzo de 2024, <https://valenciacity.es/actualidad/una-trampa-cognitiva-casi-infalible/>

31 Comodidad nihilista y conformismo tecnológico, Adolfo Plasencia, Valencia City, 30 febrero de 2024, <https://valenciacity.es/articulos-de-opinion/comodidad-nihilista-y-conformismo-tecnologico/>

llamar a eso 'nueva y más efectiva fidelización de cliente' gracias a la IA, como hacen cínicamente las *big tech*?

Las consecuencias sociales³² en cuanto a problemas de salud mental³³, se están volviendo casi una epidemia, sin que las autoridades y los legisladores fuera de EE.UU.³⁴ hayan empezado siquiera a reaccionar: Y si al *machine learning* masivo y predictivo de las plataformas lo queremos llamar 'IA', habrá que hacerlo tanto para lo bueno, como para lo malo. Así que la *IA Generativa* y predictiva que impulsa las plataformas, resulta (hay datos objetivos³⁵) que está relacionada directamente, con las causas de la actual creciente ola social de problemas de salud mental³⁶ a escala social. Pero, obviamente los evangelistas de la *IA Generativa*, como buenos creyentes y devotos, o lo relativizan, o lo van a negar por más que los apabullantes e insistentes datos objetivos estén ahí. Este estado de cosas global seguirá así mientras los legisladores no protejan a usuarios y consumidores.

Una salvedad de cara al despliegue de la regulación de la IA en Europa. Sí, la normativa europea GDPR protegió en parte los datos de los usuarios, pero no los metadatos, –la mensajería instantánea encriptada punto-a-punto, tipo Whatsapp encripta el texto de los mensajes, pero no algo con aún más valor para las plataformas: sus metadatos correspondientes que permiten

32 Gráfico del informe de FT sobre la evolución de las tasas de suicidio adolescente 1990 y 2023 en EE.UU. Y Reino Unido, <https://valenciacity.es/wp-content/uploads/2024/09/Imagen-2-486K.jpg>

33 Gráfico del informe de FT sobre la evolución reciente en salud mental de los estudiantes de secundaria de EE.UU. y UK., <https://valenciacity.es/wp-content/uploads/2024/09/Imagen-3-480-k.jpg>

34 New York City designates social media a public health hazard, The Washington Post, January 25, 2024, <https://www.washingtonpost.com/technology/2024/01/25/nyc-social-media-health-hazard-toxin/>

35 The teen mental health crisis: a reckoning for Big Tech, Jamie Smyth and Hannah Murphy. Financial Times, MARCH 26 2023, <https://www.ft.com/content/77d-o6d3e-2b9f-4d46-814f-da2646fea60c>

36 Gráfico del informe de FT sobre la evolución reciente en salud mental de los estudiantes de secundaria de EE.UU. y UK., <https://valenciacity.es/wp-content/uploads/2024/09/Imagen-3-480-k.jpg>

la hiper-segmentación—. Y me temo que con la regulación europea de la IA podrá suceder lo mismo. Dicho estado de cosas, en gran parte, tiene como consecuencia una constante erosión de la capacidad humana de decisión y de libertad individual. Y por más que la gente lo crea, estas tecnologías no son deterministas. Existen formas de uso maravillosas de esta tecnología, pero no estas que imponen las *big tech* y para explotar sus infames (para el usuario) modelos de negocio. Repito, estos usos no son inevitables. La tan mencionada IA en todas sus variantes se podría usar para mejores fines, por ejemplo, para generar prosperidad y mejorar mercados y empresas, pero no unas pocas.

No hace muchos días, impartí una conferencia titulada *Redes Sociales, Neurotargeting, Sesgos Cognitivos y Fuzzy Digital Identity*³⁷, que trataba, además del peligroso y nefasto *neurotargeting*, sobre estos temas que comento, a los alumnos y alumnas del Master de Social Media y Comunicación Corporativa en la Facultad de Dirección y Administración de Empresas (ADE) de la UPV. Es un alumnado que cursa este Master porque piensa trabajar profesionalmente en el ámbito o sector de las plataformas de redes sociales, que ya han unido su destino a la *IA Generativa*. Cuando les mencionaba que esta forma de gestionar empresas y su tecnología que las *big tech* imponen ahora no tienen en cuenta los efectos sociales perniciosos y los daños personales que están causando y les mencionaba que, en su profesión, va a ser más esencial que nunca actuar en las empresas de este campo teniendo muy en cuenta los efectos sociales y la ética, que la estética, vi más de un rostro de los asistentes a la sesión que me miraba con amplia extrañeza. No exagero.

Sigamos ahora con el despliegue de la tecnología de la *IA Generativa* y otra más de las trampas de esos usos. En cuanto al desarrollo de su tecnología, Open AI fue tramposa desde el principio. Para el entrenamiento del *Transformer* GPT-3 (Chat GPT) y el GPT-4, se apropiaron de 370.000 millones de palabras publicadas, que tenían dueño legal y/o licencias, sin pedir permiso a nadie. –Es conocido lo que dice ese mantra de Silicon Valley: «pedir perdón

37 "Redes Sociales, Neurotargeting, Sesgos Cognitivos y Fuzzy Digital Identity", Faculta ADE. UPV, <https://x.com/adolfoplasencia/status/1848639430531141640>

luego, en lugar de pedir permiso antes»-. Estas empresas tienden a eso. Pero ahora tampoco hacen lo segundo. Zuckerberg solo dijo cara a cara en el Senado a los familiares de menores fallecidos «No tenían Vds. que haber pasado por esto» ...y su empresa siguió después su camino actuando igual como si nada ocurriera. Tampoco Open AI pidió autorización a los autores ni a las editoriales más de 5 millones de libros que usaron, ni a los de miles de millones de textos y publicaciones de las que Open AI se apropió subrepticamente, e incluso de millones de textos de autor en internet con licencia, incluida Wikipedia. Eso en un mundo cuyo trabajo intelectual en ciencia, cultura, arte, o literatura está basado en la autoría humana. Pero claro, sin esa gigantesca apropiación hubiera sido imposible el modelo de negocio modelo actual de Open AI cuyo nombre desde el principio era, cínicamente, –y ahora suena desvergonzado: “IA Abierta”. Cosas que pasan. Ya hay muchas noticias de daños colaterales al respecto. Un ejemplo: el New York Times demandó³⁸ en diciembre de 2023 a Open AI y Microsoft por violar sus derechos de autor exigiendo que esta empresa tecnológica pague “los daños legales y reales” que supuso el aparente plagio de su obra en el entrenamiento de IA. La demanda también pedía destruir el *software* entrenado con sus contenidos ‘robados’. Creo que va a ser que no. Bueno, se paga la indemnización y pelillos a la mar. Todo sea porque la creencia en la IA actual siga su curso, intacta. Pero hoy ya sabemos que sobre estos hechos y otros similares está construyendo los pilares de “esta” IA y de su desarrollo posterior.

A la citada apropiación indebida a gran escala para el entrenamiento de estos *loros estocásticos*³⁹, como les llama la profesora de la Univ. de Washington Emily Bender, le han seguido las del resto de las empresas desarrolladores de LLM (*grandes modelos lingüísticos*) y de sus entrenamientos. Por ejemplo, entre otras, Nvidia, una de las más empresas más recientemente

38 The New York Times demanda a OpenAI y Microsoft por violar sus derechos de autor, Wired, 27 de diciembre 2023, <https://es.wired.com/articulos/the-new-york-times-demanda-a-openai-y-microsoft>

39 On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?, ACM, <https://dl.acm.org/doi/10.1145/3442188.3445922>

demandadas⁴⁰ legalmente, tiene ahora una batalla legal abierta con creadores de contenido literario y periodístico que se enfrenta contra el uso fraudulento que hace esta empresa de su material para el entrenar su IA. Hay mucha casuística, detalles y cuestiones alrededor de esto, desde la huelga⁴¹ del sindicato de guionistas de 2023 en Hollywood, a la indignación hace muy poco de Nicolas Cage por lo que le hicieron en su último cameo en *The Flash*, por ejemplo. Pero no hay espacio aquí para extenderme más.

No importa que la empresa Open AI que desarrolló el Chat GPT sea ahora una empresa descerebrada de su mejor inteligencia empresarial. Tengo la impresión de que eso parece que no detendrá a la nueva *religión* de la IA *Generativa*, que ya parece algo totalmente inevitable y ya posee millones de feligreses, cantidad de estómagos agradecidos, y millones de pícaras mentes divertidas y satisfechas. Las promesas de exuberancia ingente de productividad de la IA también se han convertido en otra fuerte creencia, aunque no para todo el mundo. No era una creencia verosímil para mí⁴² el año pasado, y ahora no lo es ni para Daron Acemoglu⁴³, –perdón por la osadía–. Como he dicho, hay a este respecto también, muchas más cuestiones que no caben aquí.

Mientras redacto estas palabras me llegan las noticias de última hora con un titular que yo creo que esta aquejado del que yo llamo nuevo '*sesgo cognitivo*⁴⁴ de la IA': Es del diario de más tirada en España y dice así: «Las grandes

40 Escritores demandan a Nvidia por la presunta violación de derechos de autor para entrenar IA, Wired/Ars Technica, <https://es.wired.com/articulos/escritores-demandan-a-nvidia-por-violar-derechos-de-autor-para-entrenar-ia>

41 Huelga del Sindicato de Guionistas de Estados Unidos de 2023, Wikipedia, https://es.wikipedia.org/wiki/Huelga_del_Sindicato_de_Guionistas_de_Estados_Unidos_de_2023

42 La aparente y probablemente falaz extrema productividad de la IA Generativa, Adolfo Plasencia, 28 de agosto 2023, <https://adolfoplasencia.es/blog/la-aparente-y-probablemente-falaz-extrema-productividad-de-la-ia-generativa/>

43 What if the A.I. Boosters Are Wrong?, NYT, 13 de jul. de 2024, <https://www.nytimes.com/2024/07/13/business/dealbook/ai-productivity.html>

44 El poderoso y a veces nefasto sesgo de confirmación, Adolfo Plasencia, Valencia City, 20 octubre 2024, <https://valenciacity.es/articulos-de-opinion/el-poderoso-y-a-veces-nefasto-sesgo-de-confirmacion/>

tecnológicas baten récords de ingresos y beneficios por la IA»⁴⁵. En el texto se citan los beneficios de Alphabet, Apple, Amazon, Meta y Microsoft (no menciona siquiera a Open AI). Dichos beneficios, ¿son porque han mejorado su producción, gestión, sus servicios o su tecnología realmente? Opino que no, entre las razones de este sector de 'beneficios gracias a la IA' menciona que Meta ha mejorado sus algoritmos de IA de sus redes para fidelizar a la audiencia (en formas como la que he citado antes) y mejorar su monetización. Así ha mejorado este trimestre sus beneficios en un 35,5%. Por su parte, Google, la antigua empresa de la frontera de las tecnologías de búsqueda, que ahora se dedica al negocio publicitario *global* en cuerpo y alma, y al *hardware* (los *data center*) también ha aumentado sus beneficios gracias a la IA. ...Tal como decía Paco Umbral, «...y todo en este plan.»

O sea que, según esta visión, las virtudes de la *IA Generativa* consisten, sobre todo, 'aumentar la fidelización' del usuario (mediante más manipulación personalizada en modo monopolio creciente); automatizar la gestión y 'mejorar la monetización', –algo que es un lugar común–, e imagino que eso incluye evadir impuestos e implantar sedes fiscales en las Islas Salomón como ha hecho TikTok⁴⁶ –ByteDance fuera de China–, que ya ha conseguido en geografía del gigante asiático y fuera sus primeros 1.000 millones de usuarios. Esta plataforma de propiedad china no se menciona en el artículo citado. Igual es que piensan que no usa IA, cosa que dudo.

8. LA IA GENERATIVA INJUSTA Y LA IA GENERATIVA DE DOBLE MORAL

A veces en los discursos del *marketing* rampante de la industria de las plataformas globales incluyen frases sobre la ética, pero es pura retórica hueca. Ya que he citado a ByteDance la empresa matriz de Tik Tok y Douyin (la

45 Las grandes tecnológicas baten récords de ingresos y beneficios por la inteligencia artificial, Miguel Jiménez, Washington, El País, 2 de noviembre de 2024, <https://elpais.com/economia/2024-11-02/las-grandes-tecnologicas-baten-records-de-ingresos-y-beneficios-por-la-inteligencia-artificial.html>

46 TikTok: tenemos un problema social y humano con vuestros algoritmos, Adolfo Plascencia, Valencia City, 10 octubre de 2024, <https://valenciacity.es/actualidad/tiktok-tenemos-un-problema-social-y-humano-con-vuestros-algoritmos/>

red social de TikTok, dentro de China). Veamos. Dicha empresa que usa la *IA Generativa* a gran escala es un ejemplo diáfano de lo que yo llamo la *IA Generativa Injusta* y también de *IA Generativa de Doble Moral*. Trataré de resumirlo. Para empezar, la primera está causando entre sus usuarios de todo el mundo un incremento de los problemas de adicción a gran escala con la estadística predictiva y su *scroll* infinito. La algoritmia de Tik Tok es la más rápida generando adicción intensa sobre todo en adolescentes y preadolescentes, –pero no solo–, a los que causan con diversas combinaciones de elementos tecnológicos y cognitivos desde severos y diversos problemas⁴⁷ de salud mental (ver detalle en el gráfico) a gran escala, hasta trastornos de conducta alimentaria, anorexia, e incluso hasta muertes no buscadas como ha descrito la sentencia de Juez Paul Matey, así como autolesiones, que lleguen al suicidio⁴⁸ en niños y adolescentes. Como describo en mi informe de investigación publicado⁴⁹ sobre ello, ya no son solo estadísticas y encuestas. Hay muchos datos objetivos como los de las investigaciones de la prestigiosa Pew Research, pero también ya hay una sentencia que es histórica. Intentaré resumir. Como digo en mi informe:

El Juez Paul Matey, del tribunal federal del tercer circuito (tribunal intermedio en EE.UU. entre los tribunales federales y el Supremo), ha tomado una decisión histórica contradiciendo con su sentencia⁵⁰ al tribunal anterior de primera instancia, declarando responsables a la empresa TikTok Inc. y a su matriz china ByteDance, empresas dueñas de la plataforma TikTok, de la muerte de Nylah Anderson una niña de diez años que participó en la red social en el «reto del

47 Social media companies are fueling a mental health crisis, especially for our young people. But we won't let Big Tech endanger our kids, Mayor Eric Adams @NYCMayor, <https://twitter.com/NYCMayor/status/1750227687581307123>

48 <https://valenciacity.es/wp-content/uploads/2024/09/Imagen-3-480-k.jpg>

49 <https://valenciacity.es/actualidad/tiktok-tenemos-un-problema-social-y-humano-con-vuestros-algoritmos/>

50 UNITED STATES COURT OF APPEALS FOR THE THIRD CIRCUIT No. 22-3061 <https://cases.justia.com/federal/appellate-courts/ca3/22-3061/22-3061-2024-08-27.pdf?ts=1724792413>

apagón», un horroroso e irracional reto, que le acabó costando la vida a la niña.

La información citada me llegó por Matt Stoller director de Investigación del American Economic Liberties Project y autor del libro *Goliat* que trata de la guerra de los 100 años entre el poder monopolístico y la democracia. Fue Matt quien me señaló primero la importancia histórica de esta sentencia y del juez que ha dictaminado que TikTok debe ser juzgada por manipular a niños para que se hagan daño, saltándose los anteriores criterios de aplicación de la ley durante años conocida como la Sección 230 de la Ley de Decencia en las Comunicaciones de EE.UU. Ley que concedió de facto una larga impunidad a las empresas de plataformas de redes sociales sobre los daños que causen, siempre y cuando pudieran decir que «el algoritmo fue quien lo hizo». Una sentencia valiente la de este juez, que hace retroceder la Sección 230 y pone fin al escudo sobre la responsabilidad que las grandes empresas tecnológicas utilizan para cometer malas acciones sin consecuencias y con casi total impunidad.

Las reacciones no se han hecho esperar. Por ejemplo, la profesora y jurista Zephyr Teachout de la Escuela de Leyes de la Fordham University de Nueva York, ha manifestado que «La sentencia es una decisión realmente sustancial e importante, que por fin interpreta la directiva de la Sección 230 tal y como estaba concebida, y no amplía la inmunidad a los contenidos promovidos algorítmicamente». En el mismo sentido su colega Mike Sacks señala que «el tribunal CA3 afirma en la sentencia que la Directiva. 230 no protege a TikTok y ByteDance por crear un algoritmo que alimentó el 'desafío blackout -apagón-' y tuvo como víctima directa a una niña de 10 años que intentó seguir este irracional reto hasta las últimas consecuencias y se ahorcó accidentalmente».

La Sección 230 de la Ley de Decencia en las Comunicaciones ha sido un puntal en los modelos de negocio de las actuales grandes tecnológicas y plataformas de redes sociales, dotándoles de impunidad legal desde hace mucho tiempo, a pesar de que ya es un clamor el enorme daño en salud mental que causan sus plataformas. y que

ha hecho que el mismísimo Ayuntamiento de Nueva York haya declarado⁵¹ que las redes sociales son un «peligro para la salud pública y una toxina medioambiental».

Y los daños no solo son de adicción, sino también de diversos tipos de problemas de salud mental que, en el caso de algunas edades adolescentes, se ha convertido en un auténtico problema social. En el mismo informe de Financial Times (FT) se puede ver el gráfico de la evolución estadística de suicidios infantiles y adolescentes tanto en chicos como en chicas en los últimos 30 años en EE.UU. y Reino Unido, tienen un punto de inflexión al alza, justo el año 2008, –año del nacimiento a la popularidad de este tipo de plataformas de red social y de la explosión de Facebook–. En este año, cambió la tendencia y se aceleró en el número de casos:

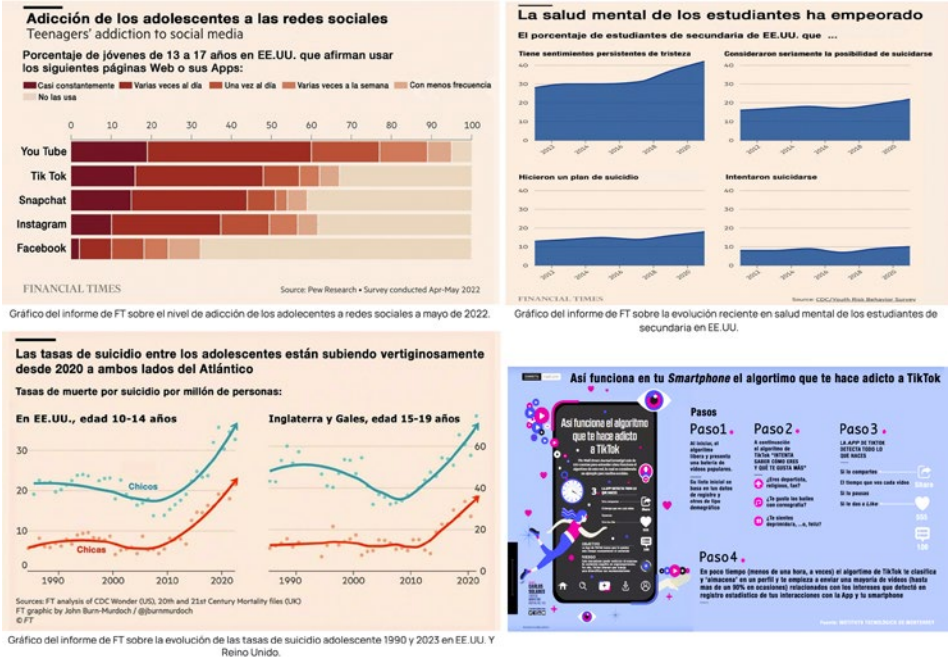
Muchos padres y legisladores culpan, y con razón, a las empresas de redes sociales que, según ellos, desarrollan productos altamente adictivos que exponen a los jóvenes a material nocivo con, a veces, fatales consecuencias en sus vidas en el mundo real. Resulta irónico que las plataformas se defiendan alegando que su tecnología permite a las personas establecer relaciones beneficiosas para la salud mental. Lo cual, no siempre es cierto. A veces sucede lo contrario.

Jonathan Haidt, psicólogo social y profesor de la Escuela de Negocios Stern de la Universidad de Nueva York declaró a FT que «Cuando aparecieron las redes sociales o Internet de alta velocidad, distintas investigaciones se encontraron, casi todas, con la misma historia. La de que la salud mental está cayendo en picado, especialmente en el caso de las chicas». Sin embargo, algunos académicos señalan que cada vez hay más estudios que son difíciles de ignorar. La proliferación de teléfonos inteligentes combinados con Internet de alta velocidad y aplicaciones de redes sociales, están reconfigurando

51 Mayor Adams Announces Lawsuit Against Social Media Companies Fueling Nationwide Youth Mental Health Crisis, City of New York, February 14, 2024, <https://www.nyc.gov/office-of-the-mayor/news/125-24/mayor-adams-lawsuit-against-social-media-companies-fueling-nationwide-youth-mental-health#/o>

el cerebro de los niños y provocando un aumento de los trastornos alimentarios, depresión y ansiedad. El informe de FT señalaba casos personales con la colaboración y permiso de las familias para concienciar a la sociedad.

No solo los algoritmos de TikTok están causando daños. También los de las otras plataformas sociales que funcionan igual, lo hacen. Una investigación de 2022 de Financial Times (FT) lo señalaba en detalle. Los gráficos a continuación, hablan por sí mismos:



Como se ve en los gráficos, los daños no solo son de adicción, sino también de diversos tipos de problemas de salud mental que, en el caso de algunas edades adolescentes, se ha convertido en un auténtico problema social. En ese mismo grafico del informe de FT se puede ver la evolución estadística de suicidios infantiles y adolescentes tanto en chicos como en chicas en los últimos 30 años en EE.UU. y Reino Unido, tienen un punto de inflexión al alza, justo el año 2008, –año del nacimiento a la popularidad de este tipo de plataformas de red social y de la explosión de Facebook–. En este año, cambió la

tendencia y se aceleró en el número de casos. El gráfico de abajo a la derecha de la figura, el algoritmo de TikTok es el que más rápido de todas las plataformas de red social más conocidas consigue más rápidamente una adicción de alta intensidad. El prestigioso The Wall Street Journal (WSJ) investigó ya en 2021 más de 100 cuentas de Tik Tok para entender cómo funciona el algoritmo de esta red, que ya se está considerando como un ejemplo puntero cuyas capacidades de *engagement* (enganche) son un ejemplo para las distintas Apps con las que interactúan sobre todo los adolescentes.

El algoritmo de TikTok es altamente personalizable y alimenta el *feed* (una banda vertical de contenido continuo y sin fin), que aparece la interfaz de la App Tiktok en pantalla del móvil al elegir *For You Page* (para tu página). Técnicamente esto se llama *scroll* infinito y fue creado para ofrecer a los usuarios un flujo interminable de contenido, priorizado a partir de lo que se ha registrado mediante la *Aptigraphy*⁵² y la *Typography* por los sensores de su Smartphone. Nada que aparece en la pantalla de la App es casual ni aleatorio como suele creer el usuario. El objetivo es asociar lo que aparece en pantalla coincidiendo con picos de emoción en la estadística de uso en el tiempo, de ese usuario en particular. Y ese tipo de contenido es el que prioriza el algoritmo para enviarlo constantemente, con el fin de que no tengas tiempo de tomar ninguna decisión, ni haya tiempos muertos e interactúes al instante, compulsivamente, sin pensar, ayudándose de las notificaciones constantes sonoras y personalizadas por la algorítmica predictiva. Esto hace que la App sea no solamente entretenida, sino que sea particularmente adictiva. Es una sofisticada herramienta ante la que los usuarios están indefensos, ya que les bombardea sin tiempo para decidir, desde que, enganchados, le siguen el juego al algoritmo.

En cuanto a la **IA Generativa de Doble Moral**, el ejemplo más evidente está en la misma empresa: el conglomerado propiedad de ByteDance que forman las plataformas de Douyin (el 'Tik Tok, chino' que opera dentro de China),

52 *Actigrafía y Tappigrafía*: tu Smartphone sabe cómo te sientes antes que tú, Adolfo Plasencia, Valencia City, 6 de octubre, 2023, <https://valenciacity.es/actualidad/tecnologiату-smartphone-sabe-como-te-sientes-antes-que-tu/>

más la plataforma de red social Tik Tok Internacional para el resto de mundo en el exterior de China. Funcionan ambas impulsadas por el mismo tipo de *algorítmica predictiva*, pero con algoritmos distintos para dentro y fuera de China. Es una especie muy sorprendente de 'doble moral' política, dobles reglas geopolíticas, y distintos algoritmos de IA para dentro que para fuera de China, por más que son las dos plataformas propiedad del mismo gigantesco conglomerado empresarial de propiedad china.

En resumen, se puede describir así:

El TikTok que se usa en todo el mundo, excepto en la geografía china, lo gestiona la empresa china TikTok inc. que tiene su sede fiscal en las Islas Salomón, un paraíso fiscal. Pero su versión en chino, –y esto es importante–, que se llama Douyin, y es la versión de TikTok de la misma empresa, que solo está disponible para usuarios del interior de China y que usa un algoritmo diferente, aunque ambas pertenecen al gigante empresarial ByteDance, con sede en China y en cuyo consejo se sienta un miembro del Comité Central del Partido Comunista de China. Dicho conglomerado, la empresa Bytedance funciona bajo la supervisión y regulación del Gobierno de China, al que entregará datos de los usuarios, incluidos adolescentes de todo el mundo ante cualquier petición de su gobierno «si fuera necesario». Aunque TikTok y Douyin comparten estética, ofrecen contenidos con métodos distintos. La versión para dentro de China de esta red social sí tiene salvaguardias y limitadores de tiempo para evitar el exceso de uso sobre todo en niños y pre-adolescentes. Hasta las autoridades chinas conocen el daño que un mal uso por sobre-exposición de las plataformas pueden hacer. De ahí que impongan severas restricciones. Cualquier padre cuyos hijos infrinjan estas limitaciones puede acabar en la cárcel ipso facto por ello.

Por lo visto, a los gestores del TikTok internacional, –el que conocemos por ejemplo en España–, en cambio, los niños y adolescentes occidentales no les parecen dignos de protección. Al contrario, en cambio, para el uso por niños y niñas chinos imponen todo tipo de salvaguardias y restricciones. Por ejemplo: para evitar la

sobreexposición, en la red social Douyin está establecida por defecto un tiempo máximo de 40 minutos en la App para los usuarios chinos menores de 14 años. También para el interior de China, el tiempo de uso para adolescentes entre los 15 y los 17 años se reduce a una hora. Y a los niños, niñas y adolescentes en China se les impide y no pueden usar Douyin de 22:00 a 06:00 horas. Además, las temáticas de Douyin están muy limitadas y deben cumplir con las políticas emanadas del Partido Comunista Chino. Lo cual no digo que sea tranquilizador.

Además, en esta red en China se promueven, como era obvio, contenidos que impulsan valores nacionales y mensajes oficiales del estado. ¿Importa todo esto a los usuarios del mundo y de España, y a los periodistas españoles que hablan y usan TikTok? Parece que no, en absoluto. Es una moda digital y eso parece ser lo único importante para los nihilistas y social media victims.

Un asunto que vale la pena señalar es que a pesar que hay sentencias como la del Juez Matey que demuestran que el uso de la algorítmica que promueven las plataformas están causando muertes, disfunciones de conducta alimentaria, y numerosos problemas de salud mental como muestran las investigaciones de Financial Times (FT) y de WSJ, y las declaraciones oficiales del Ayuntamiento de Nueva York, en la práctica no hay alarma social alguna sobre estas noticias en la sociedad, ni en los usuarios o en sus familias, hasta el punto que incluso los medios públicos, los periodistas y las universidades públicas; así como las instituciones gubernamentales, promueven el uso de Tik Tok, por ejemplo, en España. En este link está el informe en detalle sobre Tik Tok y los efectos sobre los usuarios que causa su *IA Generativa Injusta* y la *IA Generativa de Doble Moral*: <https://valenciacity.es/actualidad/tiktok-tenemos-un-problema-social-y-humano-con-nuestros-algoritmos/>



Yo creo que no se pueden seguir usando estas plataformas tan alegremente ignorando todo eso que está pasando y relativizándolo todo. Hay muchísimas cuestiones de salud, salud mental, y vidas en juego, de gente vulnerable. Es verdad que la IA parece imparable, pero no que no debería ser imparable para ciertos usos que ponen en peligro las vidas y las mentes de mucha gente. Ninguna tecnología, por fantástica que sea, se debería poner en uso masivo de la gente, sin tratar de eliminar sus efectos dañinos, ni hacerlo a cualquier precio.

Y me pregunto, ¿son estos tipos de virtudes y de usos de la *IA Generativa* que van a mejorar nuestra sociedad y la vida de los ciudadanos, que son usuarios en masa. ¿Mejorarán este tipo de *IA Generativa* nuestras Pymes, que son el 94% del total de nuestras empresas en España? Pues no lo acabo de ver. De no ser que se conviertan como los *influencers* de Youtube en simples 'comisionistas' de las plataformas y las *big tech*.

Aquello de innovar antes omnipresente, pero el término, ha pasado a mejor vida, al parecer. Tendrán que explicármelo en detalle los nuevos y emergentes expertos en esta disciplina.

9. LA IA GENERATIVA EN ESTE LADO, QUE ES EL MÍO

Pasemos ahora al otro lado, al de los usuarios. Una vez puesta a *IA Generativa* al acceso de la gente corriente, las trampas y picaresca han invadido ahora, además, también el lado de los usuarios. Desde los alumnos pícaros e inocentes que pasan al profesor como propios los escritos que le han pedido al Modelo lingüístico, a quienes firman con todo desparpajo y sin ninguna reserva, convencidos de que los textos escupidos por el Chat GPT, en realidad son suyos, aduciendo que simplemente usan el gigantesco modelo lingüístico a que le piden que realice trabajos 'intelectuales', porque según esta posición interesada 'es' simplemente un artefacto 'técnico' de apoyo y/o de ayuda, e insinuando que es una herramienta informática más, como tantas otras. Algo similar a las aplicaciones informáticas de sus PC como el tratamiento de textos MS Word, o la hoja de cálculo Excel. Pero no es lo mismo. Por ejemplo,

en gasto energético. Como acaba de publicar⁵³, Miguel Ángel García Vega, las búsquedas en Chat GPT consumen 10 veces más energía que las de Google. y el empleo de inteligencia artificial (IA) es un océano energético sin fondo. Por comparar, el gasto de energía de Chat GPT para la empresa que da el servicio a los usuarios equivale aprox. a 7.000 kilovatios/hora, por usuario, –ya que tiene detrás un data center o más–, frente a los 20 vatios/hora de energía con que funciona el cerebro humano del usuario que le escribe o le dicta los *prompt* al Chat.

Así que, de pronto, ha sucedido un milagro. En poco tiempo, y como por ensalmo, un nutrido corpus de nuevos especialistas y empresas relacionadas con la IA se ha generado y multiplicado exponencialmente, junto a millones de interesados en la nueva disciplina de moda como, si fuera una solución, tipo mágico maná, para todo tipo de problemas, mercados y empresas, algo que contradice la vocación fuertemente monopolista de las empresas de *IA Generativa*. Por otra parte, con el Chat GPT de por medio, la autoría intelectual humana genuina se ha vuelto, *fuzzy*, borrosa. Incluso hay casos en que se ha convertido en una entelequia. O bien, acaba convertida en algo no relevante, –no importa el quién, sino el resultado–. Veremos qué hacen las SGAE correspondientes con esto. Resulta que con la IA cualquiera lo puede hacer, no importa su coeficiente intelectual. Puro '*solucionismo de IA*' del que hablaba Polonski. Pero en un mundo cuyo trabajo intelectual en ciencia, en arte, o literatura está basada en el respeto a la autoría, ¿Qué puede pasar? ¿Habrà impunidad ilimitada para quienes suplanten con la IA Generativa la autoría y no respeten los métodos de los contenidos con derechos de autor? Por ahora no sabemos. Pero esto, creo, va mucho más allá de pleitos e indemnizaciones.

Por otra parte, nadie está capacitado, –porque es imposible–, para describir la trazabilidad de los contenidos desde cómo entraron en un LLM, qué pasa allí con ellos, y cómo sale lo que sale de esas mastodónticas *cajas negras* que son el Chat GPT y sus primos-hermanos, o sea, los otros modelos

53 Cara y cruz de la inteligencia artificial en la transición ambiental, Miguel Ángel García Vega, El País, 10 de noviembre 2024, <https://elpais.com/extra/energias-renovables/2024-11-10/cara-y-cruz-de-la-inteligencia-artificial-en-la-transicion-ambiental.html>

LLM (*Large Language Models*). Y, mucho menos el cómo se escupen lo que entregan esas cajas negras. Porque Chat GPT no redacta, ni escribe, –no antropoformicemos–, sino que lo que hace es, como se dice en inglés, *spits out* (escupe fuera) las palabras, igual que otros modelos generadores de imágenes y vídeos no dibujan, ni pintan imágenes, o vídeo, sino también los ‘escupen’ fuera de su *caja negra*.

Una salvedad adicional. En el caso de los grandes modelos lingüísticos LLM, se trata de complejísimas *cajas negras*, en las que los mismos que las han creado y desarrollados no saben qué sucede exactamente en su interior y, entre otros aspectos, tampoco saben cómo hacen lo que hacen. Es interesante comentar aquí, la visión de informática de la que proceden los LLM y los *Transformers GPT (Generative Pre-trained Transformers)* mascarones de proa de la popularidad de la *IA Generativa*. Según la visión citada, la dimensión y magnitud de esas *cajas negras* responden en realidad a la idea de algunos científicos mercenarios de ciertas empresas, que llegaron creo que, con las prisas, a concebir que el mecanismo profundo de “inteligencia” era cuestión de masa crítica de complejidad y capacidad de cálculo. Y, por lo visto, pensaron que aumentando la computación y la complejidad de esas *cajas negras* en unos cuantos órdenes de magnitud de iteración con respecto a las estructuras anteriores de capas de *redes neurales artificiales* podrían conseguir, que de esa masa crítica emergiera un *comportamiento inteligente*, que desembocaría a un paso de la *IA de propósito general*. Por lo que parece, no ha habido tal.

En relación a lo anterior hay publicado un magistral artículo de Will Knight en la prestigiosa *Wired* en mayo de 2017, expresivamente titulado «The Dark Secret at the Heart of AI»⁵⁴ («El oscuro secreto en el corazón de la IA») y con un subtítulo que dice: «Nadie sabe realmente cómo hacen lo que hacen los algoritmos más avanzados. Eso podría ser un problema».

En el texto se describe muy bien que existe en el núcleo íntimo de estas *cajas negras* sobre el que se han construido estas grandes estructuras de

54 The Dark Secret at the Heart of AI, Will Knight, MIT Tech Review, 11 abril, 2017, <https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/>

IA. El autor escribe: «nos se puede mirar dentro de una red neural profunda para ver cómo funciona» ...y por su naturaleza, el aprendizaje profundo es una caja negra particularmente oscura». Y añade «Las aplicaciones de IA de muchos de los servicios que se ofrece vía Web las ejecutan ordenadores que se han programado a sí mismos y lo han hecho de maneras que no podemos entender. Incluso los ingenieros que construyen estas aplicaciones no pueden explicar completamente su comportamiento». ... «Nunca habíamos construido máquinas que funciona de una forma que sus creadores no entendieran... Esto plantea preguntas alucinantes y paradójicas (*mind-blogging questions*). A medida que esta tecnología avanza, pronto cruzaremos algún umbral mas allá del cual el uso de la IA requiera un salto de fe». Y creo que no podemos seguir tranquilamente asistiendo y formando parte del despliegue global de la IA que va a afectar seguro a nuestras vidas, y seguir actuando como si todo esto no pasara, dando el citado salto de fe.

Como si fuera imposible corregir esta trayectoria, por ese camino y a partir de esa visión llegaron a desarrollar los *Large Language Model (LLM)*. Pero resulta que, además, de esa masa crítica de computación y complejidad, los LLM necesitan detrás enormes *centros de datos*, para conseguir esa inmensa tasa de computación que, por otra parte, y como era obvio, consumen una cantidad de energía gigantesca e inusitada. Unas cuentas decenas o centenares de miles de millones de dólares después, algunos se dieron cuenta que esa visión había sido un craso error, pero fueron los inversores se dieron cuenta antes de que ese camino no era económicamente sostenible, por encima de que desde el punto de vista científico también lo fuera. Como veremos después, la era de los grandes LLM de pronto la paró en seco el propio Sam Altman. Pero el sosegado y concienzudo *prueba y error* típico de la ciencia y otras tecnologías, aquí se volvió imposible por el sistema de esta tecnología, por la carrera competitiva y por la impaciencia de los inversores especuladores. La *Era de los LLM* ha sido efímera pero la inercia de ese modelo, y de variantes pequeñas continua mediante otras encarnaciones de la IA Generativa. Tanto Chomsky como Minsky advirtieron en 2011 que ese no era el camino para llegar a construir verdaderas máquinas inteligentes. El primero,

confirmando al segundo advirtió en respuesta⁵⁵ a una airada comunicación personal de Peter Norving, a la sazón, jefe de IA de Google dirigida a él, que el camino de la estadística predictiva en ese software de IA le recordaba demasiado al conductismo radical de B. F. Skinner y que, por ese camino de desarrollo tecnológico, este tipo de IA podría producir muchas ganancias económicas pero que, mediante ella, no conseguirían nunca encontrar el mecanismo profundo de la inteligencia, «porque estaban buscando en el 'dominio' y la 'categoría' equivocados.» El titular de la respuesta de Chomsky es muy ilustrativo: «Noam Chomsky on Where Artificial Intelligence Went Wrong», es decir, «Noam Chomsky sobre dónde la inteligencia artificial de equivocó.» Parece que, para algunos, o quizá para muchos, si se consigue lo primero, ya es menos importante que no se consiga lo segundo. Y siguen por ese camino. No pueden dejar de pensar que esta tecnología mejorará invariablemente.

Terminada esta reflexión parcial al margen, vuelvo al Chat GPT, la estrella mediática de este *hype*, vuelvo a mi punto de vista personal en relación a él.

Yo no me encuentro entre los que intenta que el Chat GPT les 'escupa' algo. Confieso sin acritud que no he usado ni un instante, ningún modelo o chat lingüístico ni grande ni pequeño para confeccionar este texto que está leyendo, querido lector. Tampoco ha 'escupido' ni una sola de las palabras, organizadas, estructuradas y encadenadas de este texto, ni el Chat GPT ni ninguno de los, lo reitero, 'loros estocásticos'⁵⁶ o 'Termomix' de la informática similares, como yo les suelo llamar también. Lo digo para que quede constancia por si el lector tiene dudas al respecto.

Está claro que la IA es sobre, todo en su núcleo esencial, más cosa de debatir con humanistas, ya que plantea grandes dilemas también éticos a los que ingenieros e informáticos no se pueden enfrentar sin la ayuda y las herramientas intelectuales de los humanistas; –tal vez por eso, los grandes

55 Noam Chomsky on Where Artificial Intelligence Went Wrong, Yarden Katz, The Atlantic, 1 noviembre 2012, <https://www.theatlantic.com/technology/archive/2012/11/noam-chomsky-on-where-artificial-intelligence-went-wrong/261637/>

56 Adolfo Plasencia, Workshop on Innovation on Information and Communication Technologies WIICT –2023, del Instituto ITACA de la UPV: <https://twitter.com/adolfoplasencia/status/1677044903145099265>

institutos que investigan el futuro de la inteligencia a largo plazo, en las mejores universidades del mundo, están liderados por humanistas-. Pero a mí no me sirve cualquier tipo de humanismo en esto, y mucho menos de transhumanismo. Por eso no considero alineada mi visión sobre la IA, por ejemplo, con la del filósofo sueco Nick Bostrom investigador de la Universidad de Oxford en donde no hace mucho ha dimitido de su puesto en la institución. Lo ha hecho, tras cerrar sus puertas⁵⁷ el Instituto para el Futuro de la Humanidad de Oxford (IFH) que él fundó y dirigía, que se hizo célebre por contar con financiación económica de famosos multimillonarios de Silicon Valley, como el co-fundador de Facebook Dustin Moskovitz, que donó 13,3 millones de libras; o el célebre Elon Musk, que donó un millón de libras en 2015. Con ello, se cierra, por cierto, un búnker de ideas del trashumanismo más famoso. En este caso Bostrom lo puso en marcha, para investigar hipotéticos riesgos existenciales para la humanidad, pero en cuyo fondo estaba el marco de futuras amenazas existenciales de la raza humana por el hipotético surgimiento de una super-inteligencia a partir de la actual, –según él–, imparable IA.

En conclusión, como afirma el profesor Vyacheslav Polonski también investigador de Oxford; en tan sólo unos pocos años, el *solucionismo de la IA* ha triunfado. Como él dijo ya, ha migrado de las bocas de los evangelistas de la tecnología en Silicon Valley a las mentes de los funcionarios gubernamentales y políticos de todo el mundo. Según Polonski, –lo reitero, de nuevo porque es importante–, ...el péndulo ha pasado de la noción distópica de que la IA destruirá a la humanidad a la creencia utópica de que nuestro salvador algorítmico está aquí. Y aún diré más. Ha invadido todos los eventos empresariales grandes o pequeños que se precien, gubernamentales o no, y ha generado un inmenso evangelismo comercial y oportunista en las cámaras de eco de las plataformas de ámbito empresarial y político.

57 Cierra sus puertas el Instituto para el Futuro de la Humanidad de Oxford, ZENIT, 18 junio de 2024, <https://es.zenit.org/2024/06/18/cierra-sus-puertas-el-instituto-para-el-futuro-de-la-humanidad-de-oxford/>

Mi opinión, hoy por hoy, dada la ola creciente de problemas reales y acuciantes de salud mental⁵⁸ provocada por la 'estadística predictiva' de las plataformas globales –una de las formas más nihilistas potenciada por la llamada *IA Generativa*–. Es casi ahora una epidemia social de trastornos de conducta y salud que estamos viendo crecer a nuestro alrededor entre adolescentes y no adolescentes. Creo que las autoridades deberían hacer algo al respecto ya que esta situación no debe continuar. Es fácil hablar 'políticamente' de ética en relación a la IA, pero sinceramente no la veo aplicar, fuera del *ethics washing*⁵⁹ pomposa retórica oficiosa y oficial, dado todo lo que he descrito arriba.

Ni siquiera creo que hará mella en todo esto, la ilusionante recién estrenada iniciativa de la Open Source Initiative (OSI) que ha lanzado la primera definición⁶⁰ de lo que es un *Sistema de IA de Código Abierto*. Explicarlo, en síntesis, sería un sistema de IA disponible bajo términos que otorgan a los usuarios varias libertades al modo de las que inventó para el *software libre* Richard Stallman. Según la OSI, una *IA de Código Abierto* «es un sistema de IA que se pone a disposición del usuario en condiciones y de manera que se concedan las libertades de: 1). Utilizar ese sistema para cualquier fin y sin tener que pedir permiso; 2). Estudiar cómo funciona el sistema e inspeccionar sus componentes; 3). Modificar el sistema para cualquier fin, incluso para cambiar su resultado, y 4) Compartir el sistema para que otros lo utilicen, con o sin modificaciones, y con cualquier fin». Y, además, exige a los desarrolladores de dicha IA, «proporcionar información sobre los datos de entrenamiento en cantidad suficiente para permitir que se pueda recrear el sistema de manera sustancial». Nada menos. A pesar de que igual que Henry Jenkins, me considero un utópico crítico, estoy convencido de que no será posible que la actual industria de los LLM y la *IA Generativa*, admita nunca operar bajo, o

58 Debate sobre salud mental, 8TV, 26 octubre 2024, Nos Interesa Saber, https://8tv-videos.b-cdn.net/221020241322_nis-salud-mental.mp4

59 Why Are We Failing at the Ethics of AI?, Carnegie Council, Nov 10, 2021, <https://www.carnegiecouncil.org/media/article/why-are-we-failing-at-the-ethics-of-ai>

60 The Open Source AI Definition – 1.0, Open Source Initiative, 30 octubre 2024, <https://open-source.org/ai/open-source-ai-definition>

cumpliendo con esa definición. Si hoy hay una industria tecnológica verdaderamente opaca, esa es la de la actual *IA Generativa*.

10. EL CABALLO DE LA IA, DE LA GENERATIVA Y DE LA OTRA, YA GALOPA DESBOCADO.

La otra cara de la moneda que se enfrenta a mi argumentación está en la macroeconomía tecnológica de este momento, que da pistas sobre el próximo futuro. Por mucho que queramos matizar sobre lo que se ha dado en llamar *inteligencia artificial*, es inútil negar que la economía de su infraestructura se está imponiendo ahora mismo, aunque esté en una loca carrera, –o en el camino equivocado, según Chomsky–, como he dicho antes, porque para muchos lo primero, su rentabilidad, ya es suficiente aliciente. Desde el lado macro-empresarial de esa economía, la cosa se plantea de forma completamente distinta a como yo creo que debiera considerarse. Los *factotums* económicos de la industria tecnológica dicen: seamos prácticos, de cara al próximo futuro. Lo que importan son los resultados, la hiper-productividad y las múltiples mejoras en todo tipo de empresas y servicios. Por eso, están emergiendo todo tipo de variantes, necesarias o no, de la IA, que van permear y ocupar todo el horizonte de actividades sean profesionales o de la vida personal. La IA, –a la que podrá llamar como uno quiera, a ellos les da igual–, y sus variantes va a colonizar ‘todo’ desde sus mastodónticas granjas de servidores, en las que se está invirtiendo con un furor inusitado impulsado por el acrónimo IA que en este contexto se ha convertido no solo en una creencia, sino también en una profecía auto-cumplida. El caballo de la IA, que suma la *IA Generativa* y *el resto de todas sus variantes*, ya galopa desbocado.

Según Joseph Politano, el conocido economista de la Generación Z y fundador de Apricitas Economics, en la macroeconomía global actual, la IA y sus encarnaciones ya son un factor ineludible. La demanda de IA está disparando la inversión estadounidense en ordenadores, centros de datos y otras infraestructuras físicas; los desarrolladores de la disciplina ahora compiten intensamente y cada uno apuesta por que las continuas mejoras de sus “productos” y una mayor comercialización, que va a mejorar, dicen, con creces la escala histórica de las inversiones actuales. Se espera que la inversión aumente a medida que se desarrollen Modelos más avanzados y se extienda el

uso de la IA a más aplicaciones del mundo real. Los responsables políticos del *establishment* tecnológico también ven la IA como una parte clave de la economía del próximo futuro. La visión norteamericana (y quizá la china también) es que el despliegue de la IA y la capacidad de sus centros de datos para *cloud* y plataformas son sectores punteros hoy los que EE.UU. parece haber conseguido ventajas significativas gracias al dominio de Silicon Valley y los grandes conglomerados tecnológicos estadounidenses. Creen que por ello el auge de la IA ha beneficiado a la inversión estadounidense quizá más que a ninguna otra economía. Aunque habría que preguntar a China sobre esto para ver su punto de vista.

Así que, en la macroeconomía de la IA de EE.UU., y ahora con a victoria de del tándem Trup/Musk imagino ya un nuevo slogan del tipo: "America's AI First!". Según Politano, se está viviendo un momento de entusiasmo. Esa visión sobre la citada demanda la achacan a que los llamados "productos de IA" se utilizarán en todos lados para generar código de *software*, texto e imágenes, analizar datos, automatizar tareas, optimizar plataformas *online* y todo tipo de cosas. Y los más optimistas, esperan que su uso aumente en el futuro. Sin embargo, estos Modelos de *machine learning* predictivo de vanguardia necesitan enormes recursos informáticos para su entrenamiento e inferencia. Y su enorme demanda de computación requiere conjuntos masivos de *hardware* (infraestructura avanzada alojada en grandes instalaciones con acceso a enormes cantidades de energía, agua, banda ancha y otras infraestructuras para sus operaciones).

Por eso, en este momento, la construcción de *data centers* en EE.UU. alcanza la cifra récord de 28.600 millones de dólares anuales, un 57% más que en 2023 y un 114% más que en 2022, hace dos años. Y, además, se ha producido una paradoja significativa. El aumento de la demanda de semiconductores de gama alta ha incrementado la dependencia de EE.UU. de las importaciones sobre todo de la taiwanesa TSMC, principal fabricante mundial, con Nvidia, de semiconductores de última generación. Dicha paradoja se complementa con la voraz demanda de computación de IA que se visualiza en la creciente cantidad de *chips*, servidores y componentes relacionados, que EE.UU. debe importar ahora de Taiwán. En estas informaciones del sector, ni se menciona

el cambio climático. Sus problemas asociados no parecen humanos. Justo los que me preocupan más a mí. Y esa es otra paradoja.

Como ciudadano de a pie que soy, no consigo que se me contagie esa euforia de expectativa macroeconómica de los voceros de estos nuevos señores feudales del aire, o de *La Nube*.

11. HACIA EL PRÓXIMO FUTURO DE APLICACIONES DE LAS IA

Si tomamos como modelo de la 'nueva IA' los grandes modelos lingüísticos LLM o los *Transformers* como los de la familia GPT, que se han caracterizado por, a veces, por 'alucinar'⁶¹, o sea por emitir falsedades, (una alucinación de IA se asocia a la categoría de respuestas o creencias injustificadas, aunque se les llama así por analogía con el fenómeno de la alucinación en la psicología humana) los problemas crecen.

Las *alucinaciones de la IA* que son nuevo e inesperado fenómeno que cobraron importancia en 2022, junto con el despliegue de ciertos modelos grandes de lenguaje (LLM) como Chat GPT. Los usuarios se quejaron de que estos *chat-bots* a menudo parecían incrustar "sociopáticamente" y, sin sentido, falsedades aleatorias que parecían plausibles en el contenido que generaban. En 2023, aún muchos analistas consideraban que las *alucinaciones* frecuentes eran un problema importante en la tecnología LLM. Eso es un ejemplo de un problema de estas IA en forma de *cajas negras*, que igual que el resto de su funcionamiento no es posible saber la razón exacta por la que una IA "alucina" o presenta datos falsos, a pesar de muchos análisis, para averiguar la causa; si es por el entrenamiento, es a partir de los datos o por otras razones. Da que pensar que se hayan analizado por la propia empresa desarrolladora que incluso los ha definido como: "una tendencia a inventar hechos en momentos de incertidumbre" (OpenAI, mayo de 2023); o "los errores lógicos de un modelo" (OpenAI, mayo de 2023); inventar información por completo, pero comportarse como si se tratara de hechos (CNBC, mayo de 2023); o, "inventarse la información" (The Verge, febrero de 2023).

61 Alucinación (inteligencia artificial),Wikipedia, [https://es.wikipedia.org/wiki/Alucinaci%C3%B3n_\(inteligencia_artificial\)](https://es.wikipedia.org/wiki/Alucinaci%C3%B3n_(inteligencia_artificial))

Este es un ejemplo de problemas de la *IA Generativa* que le restan algo importante a una tecnología. Como he dicho antes, citando a Rodney Brooks, parece que damos por hecho que cualquier generación de tecnología, –gracias a la inercia que la *Ley de Moore*–, rápidamente mejorará a la anterior. Pero eso en estas IA, parece que no está ocurriendo. Tal vez porque, por su propia esencia. Estas IA no ‘comprenden’ el significado de las palabras que emiten ni ‘distinguen’ entre algo verdadero y falso. (Incluso, creo que no deberíamos usar la tercera persona del plural, para referirnos a estos gigantes–cos artefactos de *software*). De esa manera, para aplicaciones que necesitan total fiabilidad estas *IA Generativa* es seguro que no sirven por ahora, si no se corrigen estos problemas de los *Modelos de IA*. Esto de cara al futuro de las múltiples aplicaciones para diferentes campos puede ser un problema muy serio.

Y la emergente industria de la IA lo ha de resolver y rápido, porque como decía el citado Joseph Politano el gran sector tecnológico están haciendo enormes inversiones porque tienen expectativas de obtener enormes beneficios económicos, seguramente inducidas por la espectacular apariencia inicial de resultados. Y esto, coreado por la seguridad total que muchísimos devotos de la nueva IA, tienen en ella, y que el Chat GPT y otros *Modelos* han conseguido que relativicen cualquier problema, gracias a su total confianza en estos artefactos. Sin embargo, creo que una carencia de la certeza de fiabilidad completa, puede ser para la industria comercial de la IA un *handicap* a gran escala, dado como he dicho antes, el caballo de la IA ya galopa desbocado.

Hay sospechas de algún problema conceptual serio en el funcionamiento de estos *Transformers* y modelos LLM. El prestigioso lingüista Noam Chomsky, y sus colegas Ian Roberts y Jeffrey Watumull explican estos problemas conceptuales en relación con el Chat GPT, como ejemplo. En un artículo publicado⁶² en el New York Times los describían así:

62 Noam Chomsky: La falsa promesa del ChatGPT, Noam Chomsky, Ian Roberts y Jeffrey Watumull, 8 marzo de 2023, <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>

ChatGPT de OpenAI, Bard de Google y Sydney de Microsoft, etc., son maravillas del aprendizaje automático. A grandes rasgos, toman enormes cantidades de datos, buscan patrones en ellos y se vuelven cada vez más competentes a la hora de generar resultados estadísticamente probables, como un lenguaje y un pensamiento de apariencia humana...

[Pero]...La mente humana no es, como ChatGPT y sus similares, un pesado motor estadístico de comparación de patrones que se atiborra de cientos de terabytes de datos y extrapola la respuesta más probable de una conversación o la respuesta más probable a una pregunta científica. Por el contrario, la mente humana es un sistema sorprendentemente eficiente e incluso elegante que funciona con pequeñas cantidades de información; no busca inferir correlaciones brutas entre puntos de datos, sino crear explicaciones...

De hecho, estos programas (de ML) están estancados en una fase prehumana o no humana de la evolución cognitiva. Su defecto más profundo es la ausencia de la capacidad más crítica de cualquier inteligencia: decir no sólo lo que es el caso, lo que fue el caso y lo que será el caso –eso es descripción y predicción–, sino también lo que no es el caso y lo que podría y no podría ser el caso. Esos son los ingredientes de la explicación, la marca de la verdadera inteligencia.

Resalto el último párrafo de estas argumentaciones en la que estos prestigiosos científicos, estas máquinas de *software* no poseen la marca de la “verdadera inteligencia”. Por eso, al principio de mi texto sugiero que no se debería usar alegremente la etiqueta semántica “inteligencia artificial” tan alegremente como si los resultados de esta tecnología generativa o fueran fruto de una *máquina inteligente* (o casi). No es una cuestión menor. Y tampoco creo que debería ser tratada como si fuera una cuestión de creencia al modo religioso.

¿Como va a ser esta nueva etapa con la IA aplicándose a todo del futuro inmediato, como auguran las grandes inversiones en curso que he descrito antes? El gran científico Niels Bohr aseguraba que «Es difícil hacer predicciones, especialmente del futuro». Tenía razón. Así que no me atrevo a hacer

una predicción exacta. Y creo que nadie puede hacerla con seguridad sobre la que IA funcione eficientemente en tantos campos al tiempo como muchos vaticinan. Pero las expectativas que han generado el *hype* parecen ilimitadas y eso pone mucha presión.

En cualquier caso, en cuanto a mis ideas actuales al respecto en relación a la IA, *Generativa*, o no, me produce preocupaciones más asociadas a pensar qué le causa esta tecnología a la gente que la usa, o si alguien la usa para modificar su conducta sin que lo sepa, y si le ocurre en ello algo justo o injusto. Vistos los resultados⁶³ de la algorítmica predictiva que ya está siendo «*potenciada con IA*».⁶⁴ Me preocupa extraordinariamente, como ilustran muchos casos muy significativos, que se esté usando esta tecnología yo diría que contra las personas, para modificar sus conductas y mentes sin que se aperciban de ello, y con preocupantes efectos secundarios.

Mi opinión relación a ello, está también en línea con las palabras más recientes del citado Daren Acemoglu, profesor del MIT y Premio Nobel de economía 2024, a Miguel Ángel García Vela en su entrevista publicada⁶⁵ el 10 de noviembre de 2024. Le dijo, y cito literalmente: «...si permitimos que Chat GPT o la IA estén controlados por unas pocas compañías tecnológicas y se conviertan en los maestros del conocimiento y silencien a los trabajadores, entonces, verdaderamente, habremos perdido el rumbo.»

Y, en cuanto a la parte humana, mi visión respecto a la IA está hoy mucho más alineada con una afirmación de la época que se hablaba de la IA como amenaza, a la que lo que sucede ahora le está dando la razón, solo que es distinta. La hizo, entre irónico y amenazante, el sabio de la tecnología Jaron Lanier, escritor, informático y compositor de música clásica estadounidense,

63 TikTok: tenemos un problema social y humano con vuestros algoritmos, Adolfo Plascencia, Valencia City, 10 octubre de 2024, <https://valenciacity.es/actualidad/tiktok-tenemos-un-problema-social-y-humano-con-vuestros-algoritmos/>

64 La inteligencia artificial (IA) en TikTok: la plataforma habla sobre su uso, Pavel Sidorenko-Bautista, Pilar Lacasa, Mitsuko Matsumoto, Infonomy, 24 mayo 2024, <https://infonomy.scimagoepi.com/index.php/infonomy/article/view/57/86>

65 Daron Acemoglu, "Si dejamos que la IA la controlen unos pocos habremos perdido el rumbo". Miguel Ángel García Vela, El País, en papel. Economía Global. 10 noviembre 2024

–quien popularizó el concepto *realidad virtual*–. Lanier declaró a The Guardian en una entrevista en marzo de 2023 que: «The danger isn't that AI destroys us. It's that it drives us insane»⁶⁶ que en una traducción libre quiere decir: «El peligro de la Inteligencia artificial no es que nos destruya, sino que nos enajene y nos vuelva locos».

Pienso que tal vez Lanier tenga bastante razón. Siga el lector con salud conviviendo con la IA, sea con la *IA Generativa* o con el resto de todas sus encarnaciones. Y ya veremos que nos depara en el futuro inmediato la cambiante IA. Toda precaución será poca. 📄

66 Tech guru Jaron Lanier: 'The danger isn't that AI destroys us. It's that it drives us insane', 3 mar 2023, <https://www.theguardian.com/technology/2023/mar/23/tech-guru-jaron-lanier-the-danger-isnt-that-ai-destroys-us-its-that-it-drives-us-insane>

07.

QUIERO
TRABAJAR
CON LA
IA.
¿POR DÓNDE
EMPIEZO?

QUIERO TRABAJAR CON LA IA, ¿POR DÓNDE EMPIEZO?

por Ignacio Despujol Zabala

Para dar los primeros pasos en el uso de la inteligencia artificial, lo mejor es que empieces por la IA generativa y, en concreto, por los chatbots de texto basados en grandes modelos de lenguaje (LLM).

ChatGPT de la empresa Openai (accesible en chat.openai.com), es el punto de partida perfecto. Su versión gratuita, accesible creando una cuenta (para lo que se puede usar la cuenta de Google), da acceso a su modelo de producción más avanzado GPT 4o, que es capaz de trabajar con texto, vídeo y audio (lo que se conoce como multimodalidad). Además, guarda los chats entre el usuario y el sistema para poder recuperarlos posteriormente.

A ChatGPT se le pregunta utilizando texto, creando lo que se conoce como prompt de entrada, para pedirle lo que queremos que haga, y es capaz de realizar multitud de tareas, desde contestar preguntas, resumir y traducir textos o proporcionar ideas o instrucciones sobre un tema, hasta explicarte cómo utilizar un programa informático determinado (como Excel, por ejemplo) o, incluso, programar por ti.

La última versión de ChatGPT gratuito incorpora funcionalidades que antes solo ofrecía la versión de pago, como son:

- el acceso a internet para buscar información de la que no dispone (pues el modelo tiene datos hasta una fecha determinada en la que acabó el entrenamiento, que, en el momento de redactar esta guía, es octubre de 2023)
- la posibilidad de subir ficheros para que los analice, la posibilidad de pedirle que trate unos datos y nos dé resultados (creando un programa para ello y ejecutándolo)
- la posibilidad de generar imágenes y audio y de analizarlos (como ejemplo, se le puede subir la foto del mecanismo de ajuste del sillín de nuestra bici y pedirle que nos diga qué hacer para bajarlo).

Incluso da acceso a unas pocas interacciones semanales con el modelo que openai acaba de sacar en el momento de redactar esta guía, el modelo o1, que incorpora capacidades de razonamiento secuencial, mejorando mucho la resolución de problemas complejos y el manejo de las matemáticas.

La versión de pago, conocida como ChatGPT Plus, vale 20 euros al mes y, teóricamente, ofrece acceso prioritario en momentos de alta utilización y posibilidad de seleccionar el modelo siempre (la versión gratis utiliza GPT 4o o GPT 4o-mini dependiendo de la carga del momento y puede bloquearse por saturación), permitiendo, además, un mayor número de interacciones semanales con el nuevo modelo o1 y dando acceso antes a las novedades. También permite interactuar totalmente por voz en la aplicación de móvil y ofrece la posibilidad de crear chatGPTs personalizados, definiendo un “prompt de sistema” (unas instrucciones sobre cómo comportarse que se superponen a lo que introduzca el usuario) y ofreciendo la posibilidad de subir unos documentos de apoyo que el sistema utilizará a la hora de contestar. Estos chatGPTs personalizados solo pueden ser creados con versiones de pago, pero pueden ser usados con cualquier versión.

Openai ha añadido una actualización muy interesante a ChatGPT plus, un nuevo modelo con una interfaz interactiva (ChatGPT 4o with canvas) que permite editar un documento o un programa a la vez que lo vas generando con la IA.

¿QUÉ ES IMPORTANTE TENER EN CUENTA AL INTERACTUAR CON UN SERVICIO DE ESTE TIPO?

Lo primero es que es necesario aprender a preguntarles, lo que se conoce en inglés como “prompt engineering”. Es muy importante hacerle preguntas claras y concisas, dándole el mayor contexto posible (no importa ser redundante). Es conveniente asignarle un rol al chatbot (“eres un experto de xxx que vas a hablar para el gran público”) y añadir ejemplos de cómo queremos la salida. Normalmente no conseguiremos el resultado buscado a la primera, con lo que es importante iterar. Dado que los modelos normales no incorporan el razonamiento, muchas veces es buena estrategia ir construyendo la respuesta final a partir de ir pidiéndole cosas al modelo poco a poco, cada

una sobre la respuesta anterior, en un modelo que emula el razonamiento humano denominado cadena de pensamientos.

Es clave tener en cuenta que los modelos se inventan cosas y las presentan de forma muy creíble (fenómeno conocido como alucinación), ya que están programados para satisfacer las consultas del usuario y su tecnología está orientada a proporcionar la palabra más probable en función del contexto, no a devolver respuestas comprobadas. Por ello debemos comprobar siempre que la respuesta que recibimos es correcta.

También es importante saber que, por defecto, los datos que se suben al hacer las preguntas se usan para entrenar el modelo, con lo que no hay que subir datos sensibles o confidenciales (casi todas las empresas están incorporando mecanismos para poder pedir que no se usen los datos subidos, incluida openai en el chatgpt gratuito, pero suele ser una opción que no está activada por defecto, hay que buscarla en las opciones de configuración).

Otra consideración a tener en cuenta es que trabajan con un número máximo de palabras a manejar en cada interacción. Esto es lo que se llama contexto, y ha ido subiendo de algo más de 4.000 tokens (un token equivale, más o menos, a 0,75 palabras) hasta los 128.000/130.000 tokens que tienen los últimos modelos estándar. Esto tokens se comparten entre el texto y ficheros de entrada (que suele requerir mucho más) y el texto generado de salida, existiendo modelos con hasta 2 millones de tokens de contexto.

También es conveniente saber que no son muy buenos resolviendo problemas que impliquen razonamiento y matemáticas.

Una vez hemos visto las consideraciones generales (que aplican a todas las interacciones con estos modelos), continuamos viendo opciones.

OTROS MODELOS GRANDES DE LENGUAJE

Tenemos los modelos de la empresa Anthropic (en <https://www.anthropic.com/>), denominados Claude, y que, en general, tienen una tasa menor de alucinaciones. La familia Claude tiene 3 modelos, Claude Haiku, el más sencillo, más rápido y menos pesado de operar, pensado para aplicaciones como el uso local en los dispositivos móviles, Claude Sonnet, el modelo intermedio, pensado para el uso diario, y Claude Opus, el modelo más



completo y pesado, diseñado para resolver problemas más complejos. Los modelos Haiku y Opus van por la versión 3 y el modelo Sonnet por la versión 3.5 en el momento de redactar esta guía. Estos modelos son multimodales (pueden analizar imágenes), pueden generar código y tienen ciertas capacidades de razonamiento. Claude 3.5 Sonnet es una muy buena opción a la hora de generar texto, ya que tiene reputación de tener pocas alucinaciones.

Otra opción es el copiloto web de Microsoft, antes conocido como Bing Chat. Microsoft ha tenido mucha relación en el desarrollo de ChatGPT, pues ha financiado con mucho dinero a OpenAI, lo que le ha permitido integrar la inteligencia artificial generativa de OpenAI en sus productos. En el buscador Bing ha integrado GPT 4o de forma gratuita (incluyendo la generación de imágenes), creando lo que en principio se llamó Bing Chat (y, por eso, se puede acceder desde <https://www.bing.com/chat> y se puede encontrar buscando "Bing Chat") y que ahora se llama Bing Copilot. Tiene la posibilidad de generar imágenes y buscar en la web y ofrece más interacciones por conversación si estás conectado a una cuenta de Microsoft. Si se usa desde el navegador Edge de Microsoft permite analizar el contenido de documentos y páginas web. Tiene el problema de que no guarda los chats.

Microsoft también ha integrado la inteligencia artificial en productos como el Office 365, donde el Copilot 365 permite generar y resumir documentos, ayuda con el Excel o genera presentaciones de Powerpoint, entre otros, el Teams, donde genera transcripciones y resúmenes en directo en las reuniones o el github Copilot, que mejora en gran medida la productividad de los programadores.

Google también ha desarrollado su familia de modelos grandes de lenguaje, que, tras pasar por varios nombres, se ha quedado con el nombre de Gemini. Disponible en versiones Ultra, Pro, Nano y Flash 1.5, es multimodal (puede generar y analizar imágenes y sonidos, incluso analizar vídeos) y se integra perfectamente en los productos de Google. La versión gratuita del chatbot de Google <https://gemini.google.com/app> utiliza Gemini Flash 1.5 con 1 millón de tokens de contexto y la de pago Gemini Pro 1.5, con 2 millones de tokens de contexto.



Google también ofrece una herramienta muy potente, llamada NotebookLLM (<https://notebooklm.google.com/>) en la que podemos crear cuadernos con documentos y enlaces web, que analiza y para los que nos prepara un resumen, nos sugiere preguntas y nos da la posibilidad de crear un índice, una cronología o una guía de estudio pulsando un botón. Ofrece también una funcionalidad muy interesante en la que dos voces sintéticas comentan el contenido de los documentos en forma de podcast en inglés.



Existe otra solución interesante, <https://poe.com>, un integrador que permite utilizar decenas de modelos de otras empresas (incluidos todos los anteriores), tanto de generación de texto como de generación de imágenes desde la misma interfaz con una sola cuenta. Tiene una versión gratuita que solo permite acceder a los modelos más simples y con un número de interacciones mensuales muy limitado, y una versión de pago que, por unos 23 euros mensuales, permite acceder a un gran número de modelos con un número mensual máximo de interacciones que se controla utilizando un sistema de puntos (las interacciones con los modelos más complejos cuestan más puntos). No ofrece búsqueda en internet, análisis de imágenes o ejecución de código, pero permite comparar la respuesta de muchos modelos muy diversos del IA de forma fácil. Para empezar, puede ser una buena opción empezar usando el chatGPT gratuito, que ofrece prácticamente la misma funcionalidad que el de pago, y comprar una licencia mensual de POE para comparar la respuesta de otros modelos.



La Universitat Politècnica de València dispone de un curso masivo abierto en línea (MOOC) gratuito de introducción al uso de ChatGPT, cuya dirección es <https://www.upvx.es/courses/course-v1:edxorg+intro-chat-gpt-es+v3/about>



Tanto Openai, como Anthropic y Google permiten el uso de sus motores de lenguaje desde Internet usando programas que los interrogan de forma programática, conectándose a ellos a través de un mecanismo denominado

API (siglas de Interfaz de programación de aplicaciones), que suele ofrecer un número pequeño de interacciones gratis al mes y luego un pago por uso de unos céntimos de euros por millón de tokens utilizados.

Además de los modelos propietarios anteriores, existen modelos grandes de lenguaje de código abierto, que pueden ser utilizados de forma gratuita si se disponen de las máquinas adecuadas para ejecutarlos (servidores que, normalmente, necesitan unos dispositivos electrónicos especiales denominados GPUs o TPUs para acelerar la ejecución de estos modelos). Hay modelos de texto, imagen, sonido y vídeo, siendo los más conocidos los de Meta (antigua Facebook), denominados Llama (Meta acaba de sacar en el momento de escribir esta guía una suite de 4 modelos Llama 3.2, dos de ellos muy ligeros para correr en dispositivos móviles y 2 más pesados con capacidad de análisis de imagen) y los de una empresa europea denominada Mistral. Poe.com proporciona acceso a estos modelos y existe una web <https://huggingface.co/> donde están disponibles para ser descargados y ejecutados en la nube, ofreciendo además conjuntos de datos para trabajar con ellos, comparativas y documentos sobre su uso.



El tamaño y potencia de los LLM se mide en billones (estadounidenses) de parámetros, y los nuevos modelos de Meta tiene 2, 3, 11 y 90 billones. Normalmente a más parámetros mejor calidad, pero han salido modelos más modernos que, con muchos menos parámetros, igualan el rendimiento de modelos más grandes anteriores.

Existe un fabricante de dispositivos electrónicos denominado Groq que fabrica dispositivos de inferencia, un tipo de tarjeta aceleradora que sirve para utilizar los modelos, pero no para entrenarlos. Estos dispositivos se caracterizan por ser más rápidos y baratos de usar que las GPUs y las TPUs. Para demostrar su capacidad Groq ha creado groq.com, que permite usar una serie de modelos de código abierto, entre los que están los de Llama y Mistral, en la nube de forma gratuita hasta un número de tokens mensual bastante alto y, si se necesitan más, se pueden comprar a un precio muy competitivo, con lo que se puede probar a desarrollar pequeñas aplicaciones.

Un desarrollo interesante son los sitios que permiten interrogar a un documento o un conjunto de ellos sobre su contenido. Sitios como <https://www.chatpdf.com/> o <https://notebooklm.google/>, que permite cargar uno o varios documentos en diversos formatos y/o páginas web y obtener resúmenes, sacar preguntas e interactuar en un chat con el contenido.



MODELOS Y HERRAMIENTAS DE GENERACIÓN DE IMAGEN, AUDIO Y VÍDEO

También existen modelos y herramientas adaptados para generar otros tipos de medios a partir de descripciones de texto.

Para generar imágenes, además de DALL-E3 (<https://openai.com/index/dall-e-3/>) , la herramienta de generación de imagen de openai integrada en chatGPT y Microsoft Copilot (donde para usarla hay que tener una cuenta de Microsoft), hay herramientas muy interesantes que permiten crear desde esquemas hasta fotos ultrarealistas, como <https://leonardo.ai/>, <https://www.midjourney.com/home>, <https://ideogram.ai/> o in-



tegradores como <https://creator.nightcafe.studio/> que permiten acceder a varios modelos (como poe.com para texto, aunque poe.com también integra varios de estos modelos de imagen). Hay también herramientas para la creación de imágenes de marketing digital como <https://www.jasper.ai/>, y las grandes suites gráficas como Adobe han incorporado la IA generativa a sus herramientas como Photoshop, Illustrator o Lightroom, además de crear una herramienta específica <https://www.adobe.com/es/products/firefly.html>



Mención aparte merece Stable Diffusion, un modelo de generación de imágenes a partir de texto de código abierto, creado por <https://stability.ai/>, del que se han creado cientos de modificaciones para aplicaciones (entre ellas las versiones de pago <https://dreamstudio.ai/> y gratuita <https://stable-diffusionweb.com/#demo> ofrecidas por la misma empresa). Un ejemplo de



personalización del modelo puede verse en <https://stable-diffusion-art.com/automatic1111/> y una modificación para crear fotos realistas en <https://civitai.com/models/4201/realistic-vision-v11>.



También hay herramientas para el tratamiento de voz, siendo la más conocida el modelo de código abierto Whisper, que permite obtener transcripciones de audio en decenas de idiomas y su traducción al inglés. Se puede probar su uso con este enlace al servicio gratuito de ejecución de código en la nube de Google llamado Colab <https://colab.research.google.com/drive/1CvvYPAFemlZdSOtgfhN541esSlZR7lc6?usp=sharing>.



Existen servicios de pago para generar voces a partir de texto en múltiples idiomas con la posibilidad de clonar voces como <https://elevenlabs.io/> (que ofrece un servicio gratuito para un número razonable de generaciones), <https://play.ht/>, <https://www.resemble.ai/> o <https://speechify.com/es/>.



Es muy interesante la aplicación experimental Audio Overviews que Google ha incluido en su herramienta NotebookLM (comentada anteriormente), y que permite subir uno o varios documentos y crear sobre ellos un podcast en inglés con dos voces sintéticas comentándolo. Se puede ver un ejemplo en <https://x.com/stevenbjohnson/status/1833968974863663229/video/1>



Hay también aplicaciones para la limpieza de audio de entrevistas para podcast como <https://podcast.adobe.com/enhance> y aplicaciones para crear canciones a partir de una descripción como <https://suno.com/>.



En cuanto al vídeo, además de modelos de texto a vídeo muy prometedores, como SORA de Openai (<https://openai.com/index/sora/>) o movie-gen de Meta (<https://ai.meta.com/research/movie-gen/>), que todavía no están ac-



cesibles al público en el momento de redactar esta guía, hay herramientas para el doblaje de vídeos a otros idiomas que incluso cambian el movimiento de los labios como <https://www.rask.ai/> y otras que generan un avatar hiperrealista a partir de un texto como <https://www.heygen.com/> (en <https://media.upv.es/player/?id=fba71afo-83a1-11ee-9406-ff64627eodo8> hay un



ejemplo de avatar con imagen y voz completamente sintéticas) o <https://www.synthesia.io/es/>. Los editores profesionales como Adobe Premier o domésticos como Capcut han incorporado también herramientas de IA.



Hay también herramientas para ayudar a los profesores en su trabajo, como <https://quizbot.ai/>, para generar actividades interactivas como <https://nolej.io/> o <https://schemely.app/> para generar lecciones. Todas son adaptaciones de las herramientas más generales de generación de texto a tareas específicas.



Y hay herramientas para ayudar al investigador en el descubrimiento de un tema, como <https://typeset.io/> o <https://www.perplexity.ai/>, en el proceso de estudio de las publicaciones existentes como <https://www.researchrabbit.ai/>, <https://consensus.app/>, <https://scite.ai/>, <https://elicit.com/> o la realización de revisiones sistemáticas <https://www.rayyan.ai/>.



Como puedes ver hay herramientas específicas para multitud de aplicaciones, además de herramientas genéricas de gran potencia. La clave está en la experimentación. Cada herramienta tiene sus fortalezas, y la mejor manera de descubrir cuál se adapta a tus necesidades es probándolas. La ganancia en productividad que puedes obtener es muy superior al tiempo que hay que invertir ¡Así que adelante, sumérgete en el fascinante mundo de la IA generativa y descubre todo lo que puede hacer por ti!

PARA IR MÁS ALLÁ

Si se quiere ir más allá de la experimentación o el uso personal, varios proveedores de computación en la nube, como son Microsoft (Azure AI y Azure Cognitive Services), Google (Vertex AI), Amazon (AWS AI Services) o IBM (Watson

X), ofrecen la posibilidad de contratar el acceso a estos modelos y a servicios auxiliares para crear aplicaciones de gran capacidad.

Para finalizar comentaremos que estas plataformas ofrecen también servicios de aprendizaje automático y agentes virtuales, adicionales a los servicios de inteligencia artificial generativa comentados anteriormente. Si se pretende explorar este mundo de forma personal se pueden utilizar aplicaciones con interfaz gráfica como <https://www.knime.com/>, <https://orange-datamining.com/>, <https://www.weka.io/> o <https://altair.com/altair-rapidminer>.



Para trabajar con aprendizaje automático también se puede aprender a utilizar las siguientes librerías de Python:

- Scikit-learn: Biblioteca fundamental para ML en Python
- TensorFlow: Biblioteca de código abierto para ML y deep learning
- PyTorch: Framework de deep learning flexible
- Keras: API de alto nivel para redes neuronales
- XGBoost: Biblioteca para gradient boosting
- LightGBM: Framework de Microsoft para gradient boosting
- Pandas: Manipulación y análisis de datos
- O de R (otro lenguaje de programación más orientado a la estadística):
- caret: Conjunto de funciones para entrenamiento y evaluación de modelos
- mlr3: Framework moderno para ML en R
- randomForest: Implementación de Random Forests
- glmnet: Modelos lineales generalizados con regularización
- xgboost: Versión R de XGBoost. 

PROYECTO KHANMIGO

Ignacio Despujol Zabala

TITULO DEL CASO DE ÉXITO	
KHANMIGO: Aplicación de la IA en la Educación: Personalización y Asistencia de Aprendizaje con Khanmigo	
DESCRIPCIÓN BREVE DEL PROYECTO	
Resumen del problema o necesidad	En el ámbito educativo, uno de los mayores desafíos es la falta de acceso a tutorías personalizadas para los estudiantes y la sobrecarga de trabajo para los profesores, lo que limita la personalización y efectividad del proceso de enseñanza-aprendizaje. La educación personalizada es conocida por mejorar el rendimiento y la comprensión de los estudiantes, pero resulta difícil de escalar en un entorno de aula tradicional.
Solución implementada con IA	Sal Khan, fundador de Khan Academy, presentó la implementación de Khanmigo, un tutor basado en inteligencia artificial que actúa como asistente para estudiantes y profesores. Khanmigo ofrece tutoría personalizada, detecta errores, guía a los estudiantes en la resolución de problemas matemáticos y facilita la comprensión de contenidos, como la codificación y literatura. Para los profesores, actúa como asistente, permitiendo la creación de planes de lecciones y proporcionando recursos adaptados a las necesidades del aula, optimizando el tiempo de preparación y mejorando la personalización de la enseñanza. Su idioma principal es el inglés, pero funciona también en español y portugués.
CONTEXTO DEL PROYECTO	
Industria o sector	Educación
Localización	Global
Duración del proyecto	En marcha desde 2023

DETALLES TÉCNICOS	
Tipo de IA Utilizada	El proyecto utiliza modelos avanzados de lenguaje natural basados en la tecnología GPT-4, que permite la interacción fluida y contextual con los estudiantes y profesores, brindando respuestas y tutorías en tiempo real.
Algoritmos o modelos	Khanmigo se basa en modelos de aprendizaje profundo, específicamente modelos de lenguaje de gran escala, que permiten no solo comprender preguntas y respuestas, sino también proporcionar tutoría personalizada y detectar errores comunes en las respuestas de los estudiantes.
Herramientas y plataformas	OpenAI GPT-4 para la implementación de la IA. Plataformas educativas de Khan Academy que integran a Khanmigo como un asistente dentro de su entorno de aprendizaje.
Datos utilizados	El sistema utiliza datos generados por las interacciones de los estudiantes con la plataforma educativa para adaptarse a las necesidades individuales, proporcionando una tutoría personalizada basada en los avances y errores de cada estudiante.
DESCRIPCIÓN DE LOS RESULTADOS OBTENIDOS	
La implementación de Khanmigo ha transformado la experiencia de aprendizaje al proporcionar a los estudiantes un tutor personalizado que mejora su comprensión de temas complejos en matemáticas, codificación y literatura. Los profesores han logrado un ahorro significativo de tiempo en la planificación de lecciones y han podido ofrecer una educación más personalizada a sus estudiantes. El uso de Khanmigo ha demostrado un aumento en el compromiso y la motivación de los estudiantes al recibir retroalimentación inmediata y adaptada a sus necesidades. Las herramientas que incorpora son: Herramientas para alumnos: ▪ <i>Companion Mode</i>: Acompaña al alumno en la plataforma de Khan Academy ofreciendo ayuda proactiva. ▪ <i>Tutor Me Math and Science y Tutor Me Humanities</i>: Guía al alumno paso a paso en problemas de ciencias, matemáticas y humanidades. ▪ <i>Practice General Subjects</i>: Proporciona problemas de práctica personalizados. ▪ <i>Help Me Focus</i>: Ofrece estrategias de enfoque y productividad. ▪ <i>Chat with Literary Figures</i>: Permite interactuar con personajes literarios. Herramientas para profesores: ▪ <i>Lesson Plan tool</i>: Genera planes de lección alineados con estándares. ▪ <i>Refresh My Knowledge</i>: Refresca conocimientos previos a la clase. ▪ <i>Exit Ticket tool</i>: Crea preguntas para evaluar la comprensión. ▪ <i>Lesson Hook tool</i>: Genera elementos para introducir lecciones. ▪ <i>Multiple-Choice Assessment tool</i>: Crea exámenes y evaluaciones	

MEJORAS EN LA EMPRESA O INSTITUCIÓN	
La integración de Khanmigo ha permitido a Khan Academy ofrecer un servicio educativo más personalizado y efectivo, consolidándose como líder en innovación educativa. Los profesores han reportado una mayor eficiencia y capacidad para enfocarse en el desarrollo de sus estudiantes, mientras que los alumnos han experimentado mejoras significativas en su rendimiento académico y participación.	
INDICADORES DE ÉXITO	
DESAFÍOS Y SOLUCIONES	
Problemas enfrentados	- Preocupaciones sobre el uso de la IA para hacer trampa. - Limitación inicial de las capacidades matemáticas de la IA.
Cómo se superaron	- Khanmigo fue diseñado para no proporcionar respuestas directas a los estudiantes, sino para guiar y fomentar el pensamiento crítico. - El equipo trabajó en mejorar la capacidad de la IA para entender y corregir errores matemáticos, permitiendo un soporte de tutoría más efectivo.
LECCIONES APRENDIDAS	
Aspectos positivos	- La importancia de un enfoque socrático en la tutoría, que fomenta la comprensión y reflexión en los estudiantes. - La necesidad de salvaguardas para evitar que la IA se convierta en una herramienta para hacer trampas.
Áreas de mejora	- Continuar adaptando la IA para entender y corregir errores complejos en diferentes disciplinas. - Aumentar la accesibilidad del sistema a más estudiantes en todo el mundo.
PROYECCIONES FUTURAS	
Planes de expansión o mejora	Khan Academy planea expandir el uso de Khanmigo a más escuelas y centros educativos, permitiendo que un mayor número de estudiantes y profesores se beneficien de esta herramienta de IA.
Posibles nuevas aplicaciones	La ampliación del uso de Khanmigo en más disciplinas y su adaptación para otros idiomas, facilitando la implementación de la IA en contextos educativos a nivel global.
CONCLUSIÓN	
Khanmigo ha demostrado el enorme potencial de la IA para revolucionar la educación, ofreciendo tutoría personalizada a cada estudiante y asistencia eficaz a los profesores. Su implementación es un claro ejemplo de cómo la tecnología puede utilizarse para potenciar el aprendizaje y mejorar los resultados educativos a escala global.	

REFERENCIAS Y RECURSOS ADICIONALES	
Enlaces a documentación y estudios de caso	Khan Academy. (2024). Our North Star is learning: Annual report. https://annualreport.khanacademy.org/ Khan, S. (April 2023). How AI could save (not destroy) education [Video]. TED Conferences. https://www.ted.com/talks/sal_khan_how_ai_could_save_not_destroy_education
Fotografías	 Fuente: Imagen generada por DALL-e (2024_11_07)  Logo de Khanmigo

PROYECTO ZG3D

José Manuel Linares

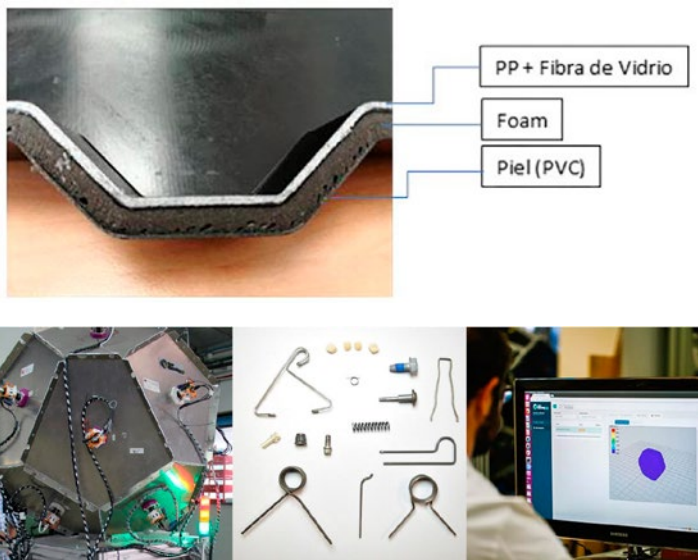
TITULO DEL CASO DE ÉXITO	
Implantación de la tecnología ZG3D para controlar contaminaciones en tiempo real.	
DESCRIPCIÓN BREVE DEL PROYECTO	
Resumen del problema o necesidad	Problema de reciclado de restos de fabricación de salpicaderos de coches debido a la contaminación del polipropileno con foam y PVC. Los clientes, como FORD y PSA, solo aceptan material reciclado si está libre de contaminantes.
Solución implementada con IA	Implantación de la tecnología ZG3D, un sistema de inspección innovador que aplica técnicas de Visión Artificial 3D en 360º para el control de calidad en línea, para controlar contaminaciones en tiempo real, en línea de proceso. Adaptación de ZG3D a medida del cliente para solucionar el problema que representa garantizar la ausencia de contaminantes en la superficie del polipropileno, lo cual habilita su reciclaje y la recuperación del coste de la materia prima.
CONTEXTO DEL PROYECTO	
Industria o sector	Inyección de plastico / Automoción
Localización	Valencia
Duración del proyecto	12 meses
DETALLES TÉCNICOS	
Tipo de IA Utilizada	Machine Learning y Deep Learning
Algoritmos o modelos	
Herramientas y plataformas	Desarrollo propio. Tecnología patentada por ITI
Datos utilizados	Entrenamiento de modelos a partir de muestras reales de producción de cada uno de los tipos de pieza, utilizando piezas ok y nok para el aprendizaje de los defectos, con etiquetado previo de defectos.

DESCRIPCIÓN DE LOS RESULTADOS OBTENIDOS	
Recuperación de materia prima por un valor de 500K€ al año con un solo dispositivo operando en dos turnos de trabajo. Además, se habilitó al cliente para reciclar garantizando la pureza del material, cumpliendo con los requisitos de las marcas fabricantes.	
MEJORAS EN LA EMPRESA O INSTITUCIÓN	
▪ Reducción de costos de producción por ahorro en materia prima ▪ Reducción de huella de carbono por reaprovechamiento de residuos que de otro modo se desperdiciaban provocando mayor consumo de materia prima y energía en su producción ▪ Aseguramiento de la calidad en el proceso ▪ Automatización de procesos	
INDICADORES DE ÉXITO	
ROI en 6 meses	
DESAFÍOS Y SOLUCIONES	
Problemas enfrentados	▪ Material de color negro, con zonas brillantes. Se tuvo que investigar en sistemas de iluminación adecuados para revelar los distintos materiales y ser capaces de detectar los contaminantes ▪ Piezas con zonas de fractura y grietas debido a la forma y al proceso de inyección, que provocaban falsos positivos ▪ Rango de tamaños de pieza variable, necesidad de un sistema capaz de funcionar en todo el rango de tamaños, desde 4cm hasta 45cm, todo con el mismo sistema sin reconfiguración ni setup ▪ Clasificación de piezas según fabricante para evitar mezclas debido a la distinta composición del material dependiendo del fabricante. Piezas muy similares y multitud de referencias distintas.
Cómo se superaron	▪ Sobre la tecnología ZG3D se ha implementado un sistema de iluminación adecuado al problema, que consigue revelar los defectos en la superficie de las piezas. ▪ Se ha trabajado intensivamente en el etiquetado de las piezas con zonas de rotura y grietas para identificar las áreas afectadas y evitar su identificación como defecto. Se ha entrenado la red neuronal especialmente para ello ▪ Una de las características principales de ZG3D es que permite el procesado en 3D y en tiempo de proceso en línea, de piezas de tamaño y formas variables sin reconfigurar la máquina entre piezas, lo que la hace muy versátil y reduce el tiempo de operación total. ▪ Se han adaptado algoritmos de clasificación usando la información 3D de cada pieza para poder identificar cada tipo de pieza y ser capaces de detectar piezas de distintos clientes y evitar así contaminación cruzada

LECCIONES APRENDIDAS	
Aspectos positivos	El proyecto ha sido un reto por tratarse de una empresa del sector automoción con estándares de calidad muy altos. El factor principal para el éxito del proyecto ha sido la estrecha colaboración con el cliente ya que la tecnología ZG3D es muy versátil y potente, pero la clave diferencial es conseguir alinearse con las necesidades particulares del cliente.
Áreas de mejora	En base al éxito obtenido, la principal mejora será poder implantar la tecnología en menos tiempo.
PROYECCIONES FUTURAS	
Planes de expansión o mejora	Ampliar la instalación a nuevas plantas de producción
Posibles nuevas aplicaciones	▪ El sistema es de aplicación directa en otras compañías que trabajan con materiales plásticos compuestos y necesitan garantizar que están limpios de contaminantes antes de su reciclado ▪ ZG3D además es un sistema de control de calidad de piezas en línea de producción, es capaz de medir según estándares de metrología industriales (GD&T), realizar análisis superficial y de detectar defectos de falta o exceso de material, siendo de aplicación en un rango amplio de procesos de fabricación

CONCLUSIÓN	
Este proyecto es un caso de éxito de transferencia de tecnología innovadora, basada en tecnología propietaria de ITI, en la que aplicamos técnicas de IA y visión artificial de última generación para automatización de procesos industriales. Ha permitido la colaboración entre la empresa cliente, ITI como centro de investigación con capacidad tecnológica para resolver problemas complejos y una ingeniería con capacidad de fabricación de maquinaria industrial para industrializar la tecnología. Es a su vez un ejemplo de las capacidades de la IA, su aplicabilidad a entornos reales y exigentes y cómo la investigación de un centro tecnológico se puede transferir para mejorar la competitividad de las empresas.	
REFERENCIAS Y RECURSOS ADICIONALES	
Enlaces a documentación y estudios de caso	https://www.zerogravity3d.com/compoundrecycler/ 
Contacto para más información	info@zerogravity3d.com
URL / QR	https://www.zerogravity3d.com/

FOTOGRAFÍA



APLICACIÓN PRÁCTICA DE LA IA EN EL ÁMBITO DE LA ECONOMÍA CIRCULAR

Victoria Majadas

TITULO DEL CASO DE ÉXITO	
Aplicación práctica de la IA en el ámbito de la Economía Circular: <i>Innovación en IA para impulsar el reciclaje en Torrelavega (Cantabria)</i>	
DESCRIPCIÓN BREVE DEL PROYECTO	
Resumen del problema o necesidad	En el marco de la estrategia de Economía Circular de la Unión Europea, se establecen objetivos ambiciosos para la recogida y tratamiento de residuos, como el aumento de las tasas de reciclaje y la reducción de residuos impropios. En España, la Ley 7/2022 de residuos y suelos contaminados exige a los ayuntamientos cumplir con medidas específicas: • Establecimiento de una nueva tasa específica, diferenciada y no deficitaria para la recogida de residuos (en abril de 2025 operativa). • Disminución de residuos generados (en 2030, un 15% menos respecto a los generados en 2010). • Recogida separada (habilitar 10 fracciones). • Aumento de las ratios de reciclaje (en 2030, el 50% de los residuos municipales deben recogerse de forma separada). • Disminución del porcentaje de impropios para cada fracción.
Solución implementada con IA	Para las fracciones de vidrio y envases ligeros, Candam Technologies implementó la solución RecySmart, un dispositivo IoT adaptable a todo tipo de contenedor urbano. RecySmart utiliza inteligencia artificial para identificar a los usuarios y caracterizar los envases depositados, como botellas de plástico, vidrio, latas y Tetra Briks, generando datos en tiempo real. Este sistema permite integrar modelos innovadores de reciclaje como RAYT ("Reward as you Throw"), SDDR Digital ("Digital Deposit Refund System") y PAYT ("Pay as you throw"). En el proyecto piloto de Torrelavega, se instalaron 20 dispositivos RecySmart en contenedores cercanos a colegios para fomentar el reciclaje a través de una competición entre alumnos, logrando así involucrar a la comunidad escolar y sus familias en la mejora de las prácticas de reciclaje.

CONTEXTO DEL PROYECTO	
Industria o sector	Sector público
Localización	Torrelavega (Cantabria)
Duración del proyecto	6 semanas
DETALLES TÉCNICOS	
Tipo de IA Utilizada	El proyecto hace uso de modelos de tipo Deep Learning. Este tipo de modelos son una subcategoría de Machine Learning y su característica principal es identificar patrones complejos en ficheros de audio, imagen y texto. El objetivo de los modelos es la clasificación de materiales usando señales acústicas como parámetro de entrada. Las señales de audio se preprocesan para facilitar la ejecución de las operaciones de las redes neuronales y reducir el coste computacional. El procesamiento tiene como entrada el audio capturado por micrófonos digitales (MEMS) y como salida coeficientes espectrales de las frecuencias de Mel (MFSC). Estos coeficientes son los parámetros de entrada del modelo de inteligencia artificial. El resultado de la ejecución de los modelos de inteligencia artificial es la probabilidad de que el evento pertenezca a una categoría determinada.
Algoritmos o modelos	La estrategia del desarrollo de los modelos siempre se ha basado en redes neuronales convolucionales (CNN) para clasificar eventos que representan el reciclado de diferentes tipos de materiales. La cantidad y tipo de eventos ha sufrido cambios a lo largo del tiempo para ajustar los modelos a los diferentes tipos de eventos que debe de clasificar. Con el objetivo de mejorar el rendimiento de los modelos y obtener un resultado más preciso, se han modificado categorías, capas de redes neuronales y tamaño de los datos de entrada. Una vez desarrollado un modelo, se emplean técnicas de optimización para integrar estos algoritmos en plataformas embebidas cuyos recursos de memoria y potencia de ejecución son muy limitados. Los modelos además deben poder ejecutarse en un tiempo determinado en las plataformas embebidas para que el usuario obtenga el resultado de la acción de reciclaje en tiempo real.

Herramientas y plataformas	• Recorder: dispositivo idéntico a la plataforma embebida final de hardware que se utiliza para capturar audio y generar los datasets que se usarán para el entrenamiento • Gitlab: infraestructura que contiene el repositorio para el desarrollo y mantenimiento del código • Python: lenguaje de programación de los modelos de inteligencia artificial • TensorFlow: entorno de trabajo específico para el desarrollo de modelos de inteligencia artificial que contiene librerías matemáticas específicas para ejecutar las operaciones de redes neuronales, carga, procesamiento y guardado de datos • Amazon Web Service: servicio en la nube para entrenar los modelos de inteligencia artificial • DVC: repositorio para guardar los datos de audio adquiridos y que se usan durante el entrenamiento • RecySmart: un dispositivo embebido que ejecuta los modelos de inteligencia artificial y clasifica los envases depositados en el contenedor
Datos utilizados	Además del desarrollo de los modelos de inteligencia artificial, es necesario obtener datos para entrenar y validar los modelos. Se han recopilado decenas de miles de muestras de audio de diferentes tipos de materiales y eventos que se han usado para el desarrollo y entrenamiento de dichos modelos.

DESCRIPCIÓN DE LOS RESULTADOS OBTENIDOS

Los resultados obtenidos en el proyecto Recysmart en Torrelavega han sido de gran impacto positivo tanto para la comunidad educativa como para el Ayuntamiento y Aguas de Torrelavega (gestor de residuos que impulsó el proyecto). Por un lado, gracias a la activa participación de los alumnos, se ha logrado inculcar una mayor conciencia ambiental entre los jóvenes, educándolos en prácticas de reciclaje responsable desde temprana edad. Para el Ayuntamiento y la empresa gestora de residuos, el proyecto ha supuesto un significativo incremento en la cantidad y calidad del material recuperado, lo que facilita el cumplimiento de las legislaciones vigentes sobre gestión de residuos. Esto, a su vez, ha contribuido a una mejora notable en la eficiencia de los procesos de reciclaje, permitiendo que más materiales sean reciclados en lugar de terminar en vertederos o ser incinerados. Además, el éxito del proyecto ha posicionado al Ayuntamiento de Torrelavega como un referente en innovación y sostenibilidad, dándole gran visibilidad y reconocimiento tanto a nivel local como regional. La reducción en los costes de vertido ha sido otro beneficio clave, ya que la menor cantidad de envases que llegan a los vertederos se traduce en ahorros económicos para la administración local.

MEJORAS EN LA EMPRESA O INSTITUCIÓN

El proyecto Recysmart ha generado varias mejoras tangibles para el Ayuntamiento de Torrelavega y Aguas de Torrelavega. La reducción de costos es uno de los logros obtenidos, ya que la mejora en la calidad del material reciclado reduce el volumen de residuos que terminan en vertederos y la tasa que se paga por ello. Además, se contribuye así a un entorno más sostenible.

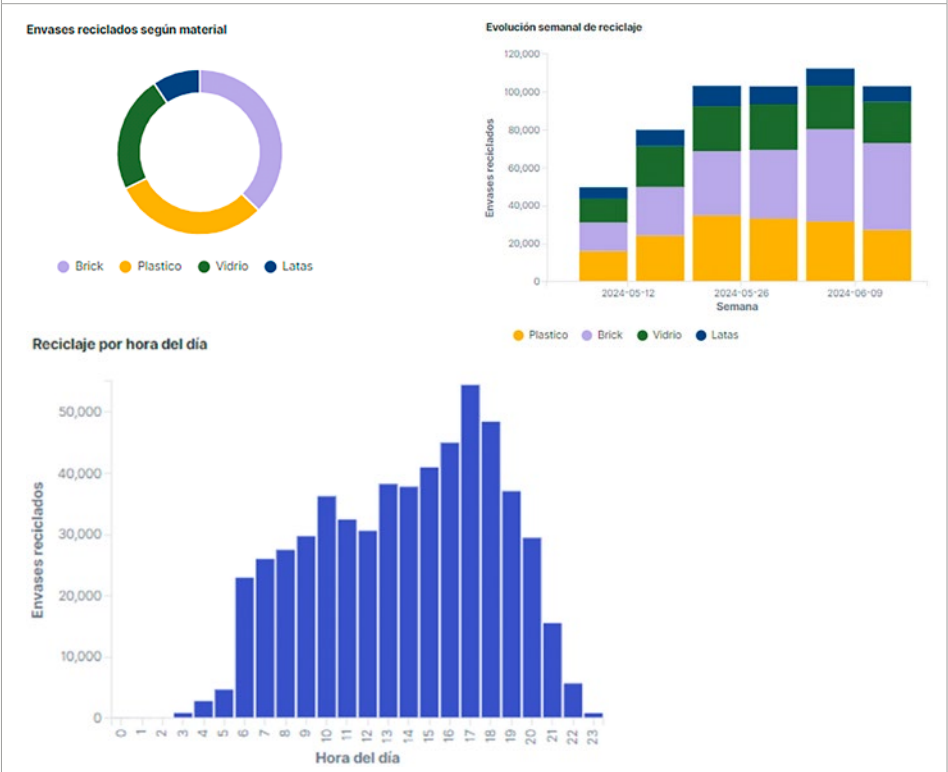
En cuanto a la satisfacción de los participantes, tanto los estudiantes como los profesores han mostrado un alto grado de compromiso y entusiasmo hacia el proyecto, lo que ha reforzado la aceptación de estas prácticas en la comunidad. La percepción positiva de la ciudadanía hacia el Ayuntamiento también ha aumentado, dado que se percibe una gestión responsable y proactiva en cuestiones medioambientales. Un aspecto clave del proyecto ha sido la capacidad de recopilar datos detallados sobre el comportamiento de los ciudadanos en cuanto al reciclaje. El sistema ha permitido identificar los horarios de mayor y menor actividad de reciclaje, los tipos de envases más reciclados, y las áreas con mejor desempeño. Esta información no solo ha sido valiosa para evaluar el impacto del proyecto, sino que también permitirá, a futuro, optimizar la recogida de residuos, mejorando aún más la eficiencia del sistema. Además, este nuevo universo de datos, inexistente con anterioridad puesto que estamos obteniendo información a nivel de contenedor y de ciudadano, contribuye a enriquecer las smart cities y al cumplimiento de ciertos objetivos de la Agenda 2030. Estas mejoras consolidan a Torrelavega como un modelo a seguir en la gestión de residuos y en la promoción de la educación ambiental, fortaleciendo su posición como un líder en innovación en el ámbito de la sostenibilidad.

INDICADORES DE ÉXITO

Algunos de los indicadores de éxito obtenidos son los siguientes:

INDICADOR	RESULTADO
Envases totales reciclados	574.911
Usuarios únicos	1.600
Participación	59%
KgCO ₂ evitados	45.826 kg
Litros de agua ahorrados	152.755 L

Finalmente, la calidad del material recolectado se puede evidenciar en el siguiente video: <https://www.youtube.com/watch?v=N5dygYFvj1o>
Otras métricas obtenidas gracias a la información recopilada en el proyecto:





Distribución horaria del reciclaje

Reciclaje por horas y días de la semana

N° Envases

0 - 2.500

2.501 - 5.000

5.001 - 7.500

7.501 - 10.000

Lunes

Martes

Miércoles

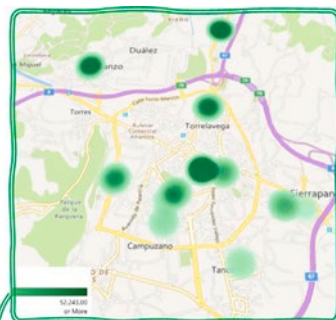
Jueves

Viernes

Sábado

Domingo

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24



Mapa de calor por zonas según número de envases reciclados

DESAFÍOS Y SOLUCIONES

Problemas enfrentados

- Necesidades de mantenimiento no planificadas: Durante la implementación del proyecto en Torrelavega, surgieron fallos técnicos en los dispositivos debido al alto uso y la participación inesperadamente elevada. No se había planificado un mantenimiento adecuado para un proyecto de estas características, lo que llevó a una dedicación de recursos de manera improvisada.
- Roturas en las tapas de los dispositivos RecySmart: En el transcurso del proyecto, se registraron varias roturas en las tapas de los dispositivos, lo que generó preocupaciones sobre su durabilidad. Aunque inicialmente se sospechaba de vandalismo, y en algunos casos se debió a ese motivo, se comprobó que las roturas ocurrieron debido a un uso intensivo de los dispositivos.

Cómo se superaron	• Mantenimiento no planificado: Se implementó un mantenimiento de emergencia con visitas semanales para solucionar los fallos. Aunque esto no fue lo más eficiente en términos de recursos, la rápida respuesta fue valorada positivamente por el cliente. • Roturas en las tapas: Se mantuvo una comunicación abierta con el cliente y se realizó un mantenimiento inmediato de las tapas dañadas. Paralelamente, se iniciaron pruebas de resistencia con robots para mejorar el diseño de las tapas en futuros proyectos.
LECCIONES APRENDIDAS	
Aspectos positivos	• Planificación de Mantenimiento: La experiencia en Torrelavega destacó la importancia de planificar el mantenimiento, incluso en proyectos de corta duración. La rápida respuesta a los problemas de mantenimiento reforzó la percepción del cliente sobre nuestro compromiso con la calidad del servicio. • Campaña de comunicación efectiva: La campaña de comunicación fue clave para el éxito del proyecto. Una estrategia bien diseñada y ejecutada resultó en una participación que superó las expectativas, con usuarios bien informados y un bajo nivel de quejas. • Identificación y solución de problemas de diseño: La detección temprana de problemas en las tapas permitió implementar soluciones rápidas y mantener la confianza del cliente.
Áreas de mejora	• Mantenimiento planificado: En futuros proyectos, es esencial incluir una planificación detallada de mantenimiento desde el inicio, independientemente de la duración del proyecto y es importante considerar sus características y posible impacto. Esto asegurará una mejor gestión de recursos y evitará problemas imprevistos. • Durabilidad de los dispositivos: Es necesario mejorar la resistencia de las tapas. Las pruebas actuales permitirán ajustar el diseño para evitar que estos problemas se repitan en futuros proyectos.
PROYECCIONES FUTURAS	
Planes de expansión o mejora	Gracias al éxito del proyecto piloto en los colegios de Torrelavega, el Ayuntamiento tiene previsto expandir la implementación de los contenedores inteligentes con la tecnología RecySmart a toda la ciudad. Esta ampliación permitirá que más ciudadanos se beneficien de la tecnología y sus recompensas, buscando seguir aumentando las tasas de reciclaje. Además, la expansión permitirá recopilar datos más completos y detallados sobre los hábitos de reciclaje a nivel municipal, lo que optimizará aún más la gestión de residuos.

Posibles nuevas aplicaciones	A futuro, el Ayuntamiento también está evaluando la posibilidad de aplicar esta tecnología en otros ámbitos, como los puntos limpios, para mejorar la eficiencia en la recogida de residuos voluminosos o especiales. Esta expansión a nuevas áreas de gestión de residuos permitirá un enfoque más integral y sostenible, facilitando la trazabilidad de los desechos y promoviendo un reciclaje más eficiente en toda la ciudad.
------------------------------	--

CONCLUSIÓN	
El proyecto Recysmart en Torrelavega ha demostrado el gran potencial de la inteligencia artificial combinada con el IoT para mejorar las tasas de reciclaje de envases. Gracias a la tecnología que convierte a los contenedores urbanos en inteligentes, con capacidad de caracterización de envases e identificación de usuarios, se logró aumentar tanto la calidad como la cantidad de envases recuperados. Además, los datos obtenidos proporcionaron información clave sobre los hábitos de reciclaje de los ciudadanos (inexistente anteriormente), lo que permitirá optimizar los procesos en el futuro. A pesar de algunos desafíos técnicos no previstos, la rápida resolución de incidencias fue bien valorada por el cliente. La alta participación de la comunidad y los resultados obtenidos consolidan a este proyecto como un caso de éxito, abriendo la posibilidad de expandirlo a mayor escala.	
REFERENCIAS Y RECURSOS ADICIONALES	
Enlaces a documentación y estudios de caso	https://candam.eu/es/actualidad/ 
Contacto para más información	Business Director – Spain: carlos.delcorte@candam.eu
URL / QR	https://www.youtube.com/watch?v=N5dyYFvj1o https://torrelavegaretorecicla.es/ https://www.eldiariotorrelavega.es/articulo/torrelavega/torrelavega-pone-marcha-proyecto-piloto-contenedores-intelligentes/20240220205351032117.html https://torrelavega.es/noticia/el-6-de-mayo-comienza-el-i-concurso-escolar-de-reciclaje-de-vidrio-y-envases-reto-reciclaje-responsable

Fotografías



AVATARES WEB CONECTADOS A CHATBOT

 Joaquín Rieta

TITULO DEL CASO DE ÉXITO	
AVATARES WEB CONECTADOS A CHATBOT Humanización de la Atención al Cliente en Web mediante Avatares 3D Interactivos y Chatbots.	
DESCRIPCIÓN BREVE DEL PROYECTO	
Resumen del problema o necesidad	La automatización de la atención al cliente en plataformas web ha llevado a experiencias poco personalizadas, lo que genera desconexión con los usuarios. La necesidad de crear una experiencia más cercana y humana en la atención virtual, que permita mantener la eficiencia de los Chatbots, ha abierto el camino hacia la implementación de avatares en 3D conectados a Chatbots inteligentes.
Solución implementada con IA	La automatización de la atención al cliente en plataformas web ha llevado a experiencias poco personalizadas, lo que genera desconexión con los usuarios. La necesidad de crear una experiencia más cercana y humana en la atención virtual, que permita mantener la eficiencia de los Chatbots, ha abierto el camino hacia la implementación de avatares en 3D conectados a Chatbots inteligentes.
CONTEXTO DEL PROYECTO	
Industria o sector	Atención al cliente digital, comercio electrónico, servicios en línea.
Localización	Global
Duración del proyecto	9 meses
DETALLES TÉCNICOS	
Tipo de IA Utilizada	Modelos de lenguaje natural (ChatGPT), reconocimiento de emociones y motores de animación 3D.
Algoritmos o modelos	Redes neuronales recurrentes (RNN) para el procesamiento del lenguaje, junto con motores gráficos para las animaciones del avatar, que incluyen Unity y Unreal Engine.

Herramientas y plataformas	Integración con APIs de OpenAI para Chatbots, ERP y CRM empresariales, y motores de animación en 3D para la generación de micro animaciones en tiempo real.
Datos utilizados	Se utilizaron grandes conjuntos de datos conversacionales y registros históricos de atención al cliente para entrenar al Chatbots, además de datos de expresiones faciales y corporales para las animaciones del avatar.

DESCRIPCIÓN DE LOS RESULTADOS OBTENIDOS	
El avatar 3D permitió mejorar la experiencia de usuario al hacer la interacción más personalizada y fluida. Los usuarios experimentaron tiempos de respuesta más rápidos y una comunicación más humana, aumentando la satisfacción en el servicio al cliente. Además, se optimizó el proceso de atención, permitiendo gestionar múltiples consultas simultáneamente a través de los sistemas conectados (ERP/CRM).	
MEJORAS EN LA EMPRESA O INSTITUCIÓN	
La empresa implementó un sistema de atención automatizado que humaniza la comunicación con los usuarios. Esto redujo los costos asociados a la atención al cliente, mientras mejoraba la retención de clientes gracias a la experiencia mejorada y el servicio en tiempo real.	
INDICADORES DE ÉXITO	
• Incremento del 45% en la tasa de retención de clientes. • Disminución del 60% en los tiempos de espera de los usuarios para recibir respuestas. • Ahorro del 50% en los costos operativos del servicio de atención al cliente.	
DESAFÍOS Y SOLUCIONES	
Problemas enfrentados	Uno de los mayores desafíos fue lograr una integración fluida entre el avatar 3D y los distintos sistemas de backend (ERP/CRM) para proporcionar respuestas precisas y en tiempo real, además de asegurar una sincronización perfecta entre las respuestas habladas del avatar y sus animaciones faciales.
Cómo se superaron	Se utilizaron APIs avanzadas y middleware para conectar los sistemas empresariales con el Chatbots, lo que permitió una transmisión de datos eficiente. Se realizaron ajustes en los motores gráficos para asegurar que las micro animaciones del avatar se sincronizaran perfectamente con las respuestas habladas del Chatbots.

LECCIONES APRENDIDAS	
Aspectos positivos	La personalización que ofreció el avatar 3D incrementó significativamente la interacción de los usuarios en la web, mejorando la satisfacción del cliente y reduciendo la necesidad de intervención humana.
Áreas de mejora	Se debe seguir trabajando en la fluidez de las animaciones faciales en escenarios de alta demanda y mejorar la calidad de las micro animaciones para hacerlas más naturales.
PROYECCIONES FUTURAS	
Planes de expansión o mejora	Se prevé seguir mejorando las capacidades conversacionales del Chatbots y agregar nuevas animaciones que reflejen emociones más complejas, además de mejorar la integración con nuevos sistemas como bases de datos internas y sistemas de facturación.
Posibles nuevas aplicaciones	Este sistema de avatar interactivo podría ser utilizado en sectores como la educación, el turismo y la atención médica para ofrecer una experiencia más inmersiva y personalizada.

CONCLUSIÓN	
El uso de avatares 3D conectados a Chatbots para humanizar la comunicación en sitios web ha mejorado la calidad de la atención al cliente, ofreciendo una experiencia interactiva, rápida y efectiva. La combinación de IA avanzada y animaciones 3D ha permitido un salto en la automatización, sin sacrificar el toque humano, lo que se traduce en una mayor satisfacción del usuario y un ahorro significativo en costos operativos.	
REFERENCIAS Y RECURSOS ADICIONALES	
Enlaces a documentación y estudios de caso	https://digitalidoso.com/producto/chatgpt-para-usuarios-de-cero-a-experto/ 
Contacto para más información	joaquin.rieta@gmail.com

INTEGRACIÓN DE CHATGPT

 Joaquín Rieta

TITULO DEL CASO DE ÉXITO	
Automatización e integración de chatGPT con google sheets y redes sociales	
DESCRIPCIÓN BREVE DEL PROYECTO	
Resumen del problema o necesidad	El proceso de creación, revisión y publicación de contenidos en redes sociales puede ser tedioso y consumir mucho tiempo, especialmente para empresas con una gran cantidad de publicaciones diarias. La necesidad de automatizar estos procesos, mientras se mantiene la calidad y coherencia del contenido, ha llevado al uso de la IA para optimizar la producción de copywriting.
Solución implementada con IA	Se implementó un sistema que integra ChatGPT para la generación automática de copies para redes sociales. El proceso comienza con una solicitud por voz, por correo o al pegar un enlace de una noticia. ChatGPT genera el texto y lo envía automáticamente a un Google Sheet para revisión. Una vez aprobado, el contenido se puede publicar directamente en las plataformas de redes sociales, todo gestionado desde una misma interfaz. Esta solución agiliza la producción de contenido y mejora la eficiencia en la gestión de redes sociales.
CONTEXTO DEL PROYECTO	
Industria o sector	Marketing digital, redes sociales, automatización de contenido.
Localización	Global
Duración del proyecto	6 meses
DETALLES TÉCNICOS	
Tipo de IA Utilizada	Procesamiento del lenguaje natural (PLN) mediante ChatGPT.
Algoritmos o modelos	Modelos de redes neuronales basados en GPT-4 para la generación de texto y comandos por voz.

Herramientas y plataformas	Integración con Google Sheets mediante API de Google, uso de herramientas de publicación de redes sociales como Buffer y Hootsuite para automatizar la publicación. Comandos de voz activados a través de asistentes virtuales como Google Assistant o Alexa.
Datos utilizados	Datos históricos de publicaciones de redes sociales, enlaces de noticias y prompts generados por los usuarios. Se utilizaron comandos por voz para automatizar los flujos de trabajo de la creación de contenido.

DESCRIPCIÓN DE LOS RESULTADOS OBTENIDOS	
El sistema de automatización ha permitido a los equipos de marketing reducir el tiempo dedicado a la creación y publicación de contenido en redes sociales. Se logró una mayor rapidez en la producción de copias de calidad, revisados fácilmente a través de Google Sheets, y la integración con herramientas de publicación permitió que el contenido estuviera en las redes sin intervención manual.	
MEJORAS EN LA EMPRESA O INSTITUCIÓN	
La empresa experimentó una mejora significativa en la eficiencia del equipo de redes sociales, optimizando el tiempo de creación y publicación de contenido en un 50%. Además, la implementación de la IA redujo errores humanos y facilitó el seguimiento del contenido antes de su publicación.	
INDICADORES DE ÉXITO	
- Reducción del 50% en el tiempo de producción de contenido. - Aumento del 40% en la eficiencia del equipo de redes sociales. - Disminución del 30% en errores de publicación gracias a la revisión previa en Google Sheets.	
DESAFÍOS Y SOLUCIONES	
Problemas enfrentados	Uno de los principales desafíos fue integrar múltiples sistemas (ChatGPT, Google Sheets y las plataformas de redes sociales) de manera fluida, y asegurar que las solicitudes por voz y correo electrónico fueran interpretadas correctamente por el sistema de IA.
Cómo se superaron	Se desarrollaron procesos específicos de validación para asegurar la correcta integración entre los sistemas. Se añadieron comandos de reconocimiento de voz avanzado y se optimizó el manejo del lenguaje natural para asegurar que las solicitudes se procesaran adecuadamente y que los copios se ajustaran a las expectativas del usuario.

LECCIONES APRENDIDAS	
Aspectos positivos	La integración de ChatGPT con Google Sheets y redes sociales demostró ser un método eficaz para reducir tiempos y optimizar procesos en la gestión de contenido. Además, la posibilidad de solicitar copias por voz o email aumentó la flexibilidad del sistema.
Áreas de mejora	Aún se pueden mejorar ciertos aspectos de personalización en los copios generados, como el tono de voz o el estilo de escritura para marcas específicas, y la integración con más plataformas de redes sociales.
PROYECCIONES FUTURAS	
Planes de expansión o mejora	El próximo paso será integrar el sistema con otras herramientas de productividad, como Trello o Slack, para facilitar aún más la colaboración en equipo y la gestión de tareas relacionadas con el contenido. También se planea implementar un sistema de retroalimentación automática para mejorar los resultados de los copios generados.
Posibles nuevas aplicaciones	Este sistema podría aplicarse en la creación automatizada de newsletters, correos electrónicos promocionales, y contenido para blogs, todo gestionado desde una sola interfaz con la misma capacidad de revisión y aprobación.

CONCLUSIÓN	
La integración de ChatGPT con Google Sheets y redes sociales ha permitido una automatización fluida del proceso de creación y publicación de contenido, optimizando el tiempo y los recursos del equipo de marketing. Las capacidades de generación de texto a través de IA, combinadas con herramientas de revisión y publicación automatizadas, han mejorado significativamente la eficiencia y calidad del contenido generado.	
REFERENCIAS Y RECURSOS ADICIONALES	
Enlaces a documentación y estudios de caso	https://digitalidoso.com/casos-de-exito/ 
Contacto para más información	Joaquin.rieta@gmail.com

AVATARES PERSONALIZADOS PARA VIDEOS EDUCATIVOS Y PUBLICITARIOS

 Joaquín Rieta

TITULO DEL CASO DE ÉXITO	
Creación de Avatares Personalizados para Videos Educativos y Publicitarios mediante IA.	
DESCRIPCIÓN BREVE DEL PROYECTO	
Resumen del problema o necesidad	En un mundo donde la personalización y la interactividad son cada vez más importantes, las empresas buscan formas eficientes de producir contenido audiovisual atractivo. La necesidad de adaptar el contenido a diferentes audiencias, idiomas y estilos visuales ha llevado a explorar la creación de avatares generados por IA para videos.
Solución implementada con IA	Se implementó una solución basada en IA para generar avatares que pueden ser personalizados según las necesidades del cliente, adaptando las expresiones faciales, gestos, y estilo visual para integrarlos en videos. Estos avatares pueden hablar múltiples idiomas y reflejar la identidad de la marca, facilitando la producción masiva de contenido sin necesidad de actores reales. La IA utiliza modelos de reconocimiento facial y transferencia de estilo para asegurar una integración fluida y realista.
CONTEXTO DEL PROYECTO	
Industria o sector	Educación, Marketing digital, producción audiovisual.
Localización	Internacional.
Duración del proyecto	6 meses
DETALLES TÉCNICOS	
Tipo de IA Utilizada	Modelos de Deep Learning y Transferencia de Estilo.
Algoritmos o modelos	Se utilizaron modelos de redes neuronales convolucionales (CNN) para el reconocimiento facial y de gestos, y GANs (Generative Adversarial Networks) para la creación de los avatares.
Herramientas y plataformas	Se implementaron en TensorFlow y PyTorch, con servidores en la nube con GPUs para la generación en tiempo real de los avatares.

Datos utilizados	Bases de datos de imágenes faciales, gestos humanos, y grabaciones de voz para entrenar los avatares en la sincronización de voz y movimiento.
------------------	--

DESCRIPCIÓN DE LOS RESULTADOS OBTENIDOS	
Los avatares generados con IA han permitido crear videos personalizados a escala con un alto grado de realismo, mejorando la velocidad de producción y reduciendo los costos en comparación con el uso de actores. Además, se logró una integración perfecta en varios idiomas, lo que permitió alcanzar a audiencias globales.	
MEJORAS EN LA EMPRESA O INSTITUCIÓN	
La solución permitió a las empresas crear contenido audiovisual de alta calidad de forma rápida y económica, mejorando su capacidad de personalización y aumentando el engagement de los usuarios. Además, la flexibilidad de los avatares permitió adaptarlos a diferentes mercados sin necesidad de nuevos rodajes.	
INDICADORES DE ÉXITO	
- Reducción del 70% en el tiempo de producción de videos. - Incremento del 35% en el engagement de los usuarios en campañas de marketing que utilizaron los avatares generados. - Ahorro del 50% en costos de producción audiovisual.	
DESAFÍOS Y SOLUCIONES	
Problemas enfrentados	Uno de los principales desafíos fue lograr una sincronización perfecta entre la voz y los movimientos faciales del avatar, además de asegurar que las expresiones faciales reflejaran emociones de forma natural.
Cómo se superaron	Se implementaron modelos avanzados de reconocimiento de emociones faciales y se realizó un ajuste fino de los movimientos de labios y expresiones mediante aprendizaje supervisado para mejorar la naturalidad y precisión de los avatares.
LECCIONES APRENDIDAS	
Aspectos positivos	La capacidad de personalización de los avatares y la flexibilidad de la solución ofrecieron grandes beneficios en términos de escalabilidad y eficiencia.
Áreas de mejora	Se identificó la necesidad de mejorar la calidad de los datos utilizados para la sincronización de gestos más complejos y expresiones emocionales.

PROYECCIONES FUTURAS	
Planes de expansión o mejora	Se planea continuar mejorando los modelos de IA para aumentar el realismo de los avatares y expandir la solución a nuevos mercados, incluyendo el sector educativo y la atención al cliente.
Posibles nuevas aplicaciones	Los avatares generados por IA podrían integrarse en plataformas de aprendizaje en línea, permitiendo a los profesores crear lecciones personalizadas y mejorar la interacción con los estudiantes.

CONCLUSIÓN	
El uso de avatares generados por IA para la creación de videos ha permitido a las empresas innovar en la producción audiovisual, reduciendo costos y tiempos, mientras que se mantienen altos niveles de personalización y engagement. Aunque existen desafíos relacionados con la naturalidad de las expresiones faciales, la tecnología sigue evolucionando y ofreciendo nuevas oportunidades para mejorar la producción de contenido a escala global.	
REFERENCIAS Y RECURSOS ADICIONALES	
Enlaces a documentación y estudios de caso	https://economistascv.org/curso-gestores-transformacion-digital-gtd/ 
Contacto para más información	joaquin.rieta@gmail.com
Fotografía	 PRESENTACIÓN DEL CURSO

DETECCIÓN AUTOMÁTICA DE CÁNCER DE MAMA

 Vicente Alepuz

TITULO DEL CASO DE ÉXITO	
Detección automática de cáncer de mama mediante aprendizaje profundo	
DESCRIPCIÓN BREVE DEL PROYECTO	
Resumen del problema o necesidad	La detección temprana del cáncer de mama es fundamental para mejorar los pronósticos y las tasas de supervivencia de los pacientes. Sin embargo, el proceso de lectura e interpretación de las mamografías por parte de radiólogos expertos presenta varios desafíos: 1. Carga de trabajo: • Los radiólogos se enfrentan a una gran cantidad de estudios de imagen que deben interpretar, lo cual puede generar fatiga y errores. • Esto puede causar retrasos en el diagnóstico y la atención oportuna a los pacientes. 2. Variabilidad en el desempeño: • Existe una considerable variabilidad en la capacidad de los radiólogos para detectar señales sutiles de cáncer en las mamografías. • Algunos radiólogos pueden tener una mayor tasa de detección que otros, lo que afecta la consistencia y la calidad del diagnóstico. 3. Falsos positivos y negativos: • La interpretación de las mamografías conlleva un riesgo de generar falsos positivos, lo que lleva a exámenes y procedimientos adicionales innecesarios. • Asimismo, existe el riesgo de falsos negativos, donde se pasa por alto la presencia de lesiones cancerosas.
Solución implementada con IA	Ante estos desafíos, surge la necesidad de desarrollar herramientas de apoyo al diagnóstico que puedan mejorar la precisión, eficiencia y consistencia en la detección temprana del cáncer de mama, complementando el trabajo de los radiólogos.
CONTEXTO DEL PROYECTO	
Industria o sector	Salud, oncología
Localización	Instituto de Investigación Oncológica de la Universidad de Michigan, Estados Unidos

Duración del proyecto	2 años (2018-2020)
DETALLES TÉCNICOS	
Tipo de IA Utilizada	Aprendizaje profundo (deep learning)
Algoritmos o modelos	Redes neuronales convolucionales (CNN) Arquitectura de red U-Net para segmentación de imágenes Transfer learning a partir de modelos pre-entrenados como VGG16 y ResNet
Herramientas y plataformas	Framework de deep learning: TensorFlow Entorno de desarrollo: Python, Jupyter Notebook GPU: NVIDIA Quadro RTX 6000
Datos utilizados	Conjunto de imágenes de mamografías de más de 70.000 pacientes Etiquetado por expertos en radiología para identificar lesiones de cáncer de mama

DESCRIPCIÓN DE LOS RESULTADOS OBTENIDOS
El modelo de IA logró una precisión del 90% en la detección de cáncer de mama en mamografías. Superó el desempeño de los radiólogos expertos, quienes alcanzaron una precisión del 85% en promedio. El modelo fue capaz de detectar lesiones cancerosas de manera más precisa y consistente que los médicos.
MEJORAS EN LA EMPRESA O INSTITUCIÓN
Permitió reducir el tiempo de diagnóstico y el número de falsos positivos. Liberó a los radiólogos de tareas rutinarias, permitiéndoles enfocarse en casos más complejos. Mejóro la calidad y eficiencia del proceso de detección de cáncer de mama.
INDICADORES DE ÉXITO
Precisión del 90% en la detección de cáncer de mama Reducción del 20% en el tiempo de lectura de mamografías Aumento del 15% en la tasa de detección temprana de cáncer de mama

DESAFÍOS Y SOLUCIONES	
Problemas enfrentados	Obtener un dataset de imágenes de mamografías de alta calidad y anotado adecuadamente. Optimizar la arquitectura y los hiperparámetros de la red neuronal.
Cómo se superaron	Colaboración con hospitales y centros médicos para recopilar un dataset amplio y de calidad. Iteración y ajuste fino del modelo a través de técnicas de transfer learning y validación cruzada.
LECCIONES APRENDIDAS	
Aspectos positivos	La calidad y representatividad del dataset de entrenamiento es crucial para el éxito del modelo de IA. La colaboración entre expertos en medicina e ingenieros de IA es fundamental para desarrollar soluciones efectivas. La IA debe ser implementada de manera complementaria al trabajo de los profesionales médicos, no para reemplazarlos.
Áreas de mejora	Ampliar el dataset de entrenamiento: Esto ayudaría a mejorar la generalización del modelo y su robustez ante la diversidad de casos. Integración en el flujo de trabajo clínico: Se podría incluir la integración con los sistemas de información hospitalarios, la interfaz de usuario, los flujos de trabajo, etc. Estudios de impacto clínico y de costo-efectividad: Analizar la relación costo-beneficio de la adopción de esta tecnología en el contexto del sistema de salud. Aspectos éticos y de privacidad: Abordar cuidadosamente los temas relacionados con la privacidad de los datos de los pacientes, el sesgo algorítmico y la transparencia en la toma de decisiones.
PROYECCIONES FUTURAS	
Planes de expansión o mejora	Integración del modelo de IA en la práctica clínica diaria de los centros de radiología.
Posibles nuevas aplicaciones	Expansión a la detección de otros tipos de cáncer mediante técnicas de aprendizaje profundo. Desarrollo de herramientas de asistencia a la toma de decisiones médicas basadas en IA.

CONCLUSIÓN

1. Potencial de la IA para mejorar la detección temprana:
 - La implementación de sistemas de aprendizaje profundo de IA se presenta como una solución prometedora para abordar los desafíos de la interpretación de mamografías por radiólogos.
 - La IA podría mejorar la precisión, eficiencia y consistencia en la detección temprana del cáncer de mama.
2. Complementariedad entre radiólogos y la IA:
 - La IA no busca reemplazar a los radiólogos, sino más bien complementar su trabajo, ayudándoles a procesar una mayor cantidad de estudios de imagen y reducir errores.
 - Esto podría permitir una atención más oportuna a los pacientes y mejorar los resultados de salud.
3. Necesidad de validación y regulación:
 - Antes de implementar sistemas de IA en entornos clínicos, sería crucial validar exhaustivamente su desempeño y seguridad.
 - Es probable que se requiera un proceso de regulación y aprobación por parte de las autoridades sanitarias correspondientes.
4. Consideraciones éticas y de equidad:
 - Al implementar soluciones de IA en salud, es importante considerar aspectos éticos, como la transparencia, la equidad en el acceso y la mitigación de sesgos algorítmicos.

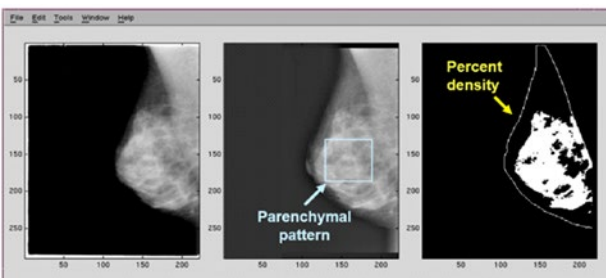
REFERENCIAS Y RECURSOS ADICIONALES

Enlaces a documentación
y estudios de caso
Contacto para más
información

[https://cad-ai.med.umich.edu/
research-projects/breast-cancer/
lesion-detection](https://cad-ai.med.umich.edu/research-projects/breast-cancer/lesion-detection)



FOTOGRAFÍA



PROYECTO BISOC

José Manuel Linares

TITULO DEL CASO DE ÉXITO	
Proyecto Bisoc. Herramienta para la búsqueda y recomendación de estrategias de marketing digital basada en IA Generativa.	
DESCRIPCIÓN BREVE DEL PROYECTO	
Resumen del problema o necesidad	El objetivo del proyecto se centra en la necesidad de optimizar la creación de campañas de marketing digital para Social Commerce. Es decir, poder acceder a contenido digital publicado en RRSS y poder recomendar y facilitar la construcción de las campañas mediante IA.
Solución implementada con IA	Se ha implementado una solución basada en técnicas de <i>Retrieval Augmented Generation (RAG)</i> y los grandes modelos multimodales (LMM) para la búsqueda, clasificación y recomendación de contenidos digitales para las campañas.
CONTEXTO DEL PROYECTO	
Industria o sector	Marketing
Localización	Valencia
Duración del proyecto	12 meses
DETALLES TÉCNICOS	
Tipo de IA Utilizada	IA Generativa basada en Large Language Models (LLM) y Large Multimodal Models (LMM)

Algoritmos o modelos	clip-ViT-B-32-multilingual-v1 - https://huggingface.co/sentence-transformers/clip-ViT-B-32-multilingual-v1 idefics2 - https://huggingface.co/HuggingFaceM4/idefics2-8b Llama-3-8B-Instruct - https://huggingface.co/meta-llama/Meta-Llama-3-8B-Instruct
Herramientas y plataformas	OpenSearch, ChromaDB, HuggingFace, Docker y CLIPScorer
Datos utilizados	Publicaciones de RRHH y Tiktok, formado principalmente por texto como imágenes y vídeo. Volumen de datos de 100K publicaciones diarias.

DESCRIPCIÓN DE LOS RESULTADOS OBTENIDOS	
El principal resultado ha sido un RAG que implementa las capacidades de búsqueda de información de contenidos digitales en RRSS basada en prompts mediante lenguaje natural o bien utilizando imágenes o vídeos. Para ello ha sido necesario implementar un clasificador de la información de los contenidos digitales capturados por la empresa, llegando a 100K de publicaciones diarias de media. Además, se le ha dotado de la capacidad de recomendar al ingeniero de campañas el tono, lenguaje y estrategia de creación de campañas de marketing.	
MEJORAS EN LA EMPRESA O INSTITUCIÓN	
Se ha habilitado una nueva funcionalidad a la plataforma de gestión de estrategias de campaña de la empresa, pudiendo así acelerar el proceso y mejorar la precisión con la que se realizaba las búsquedas y la clasificación de contenidos. En concreto, se ha pasado de una precisión del 72% a un 98% en la clasificación de las publicaciones.	
INDICADORES DE ÉXITO	
Número de publicaciones clasificadas al día.	
DESAFÍOS Y SOLUCIONES	
Problemas enfrentados	El problema no escalaba mediante servicios en la nube debido a la gran cantidad de contenidos que se deben procesar diariamente. El coste del proyecto era muy elevado.
Cómo se superaron	Se investigaron modelos opensource para la construcción de una solución local con recursos de GPU limitados.
LECCIONES APRENDIDAS	

Aspectos positivos	Existen soluciones actualmente de LLM basados en CPU que permiten abaratar los costes del procesamiento masivo en detrimento de la calidad de los resultados.
Áreas de mejora	Investigar en destilería y finetuning de grandes modelos para reducir el tamaño de los mismos y adaptar la salida a las necesidades concretas de la empresa. Además, mejorar en técnicas de ingeniería de prompts.
PROYECCIONES FUTURAS	
Planes de expansión o mejora	La creación de sistemas de recomendación para que los Influencers mejoren en sus habilidades blandas. Para ello, se emplearán técnicas de Grah Analytics.
Posibles nuevas aplicaciones	Recomendación de rutas formativas para mejorar en las capacidades de comunicación dentro de los objetivos de campañas de marketing digital.

CONCLUSIÓN	
Los sistemas RAG se han mostrado como una herramienta poderosa a la hora de emplear grandes modelos en multitud de tareas: clasificación, búsqueda, pregunta-respuesta, recomendación, generación, etc. Sus arquitecturas basadas en agentes permite además su adaptación a diferentes contextos, pudiendo emplear modelos en la nube o locales para abaratar costes.	
REFERENCIAS Y RECURSOS ADICIONALES	
Enlaces a documentación y estudios de caso	No se idsponen
Contacto para más información	Raúl Hussein Galindo
URL / QR	www.brandmanic.com 
Fotografía	

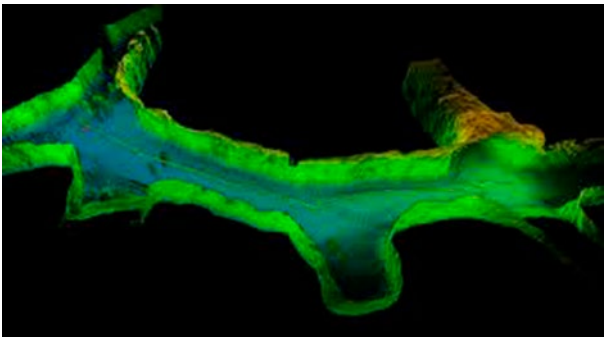
PROYECTO LIDAR

🗨 Jorge J. Bañón

TÍTULO DEL CASO DE ÉXITO	
Detección automática de derrumbamientos con tecnología LIDAR para la prevención de riesgos laborales.	
DESCRIPCIÓN BREVE DEL PROYECTO (100-150 PALABRAS)	
Resumen del problema o necesidad	Comprobación del estado de las paredes y techos de las minas con fines preventivos para la mejora de seguridad y salud del trabajador.
Solución implementada con IA	Análisis del Terreno: Detectar las paredes y techos a través de tecnologías LiDAR de forma periódica para su posterior comparación y análisis. Resultados: Obtención de las deformaciones y la velocidad a la que se desplaza la zona de interés.
CONTEXTO DEL PROYECTO	
Industria o sector	Topografía e ingeniería de minas
Localización	3274 Boul Saint-Martin O, Suite. 203, H7T 1A1, Laval, Qc, Canada Newark (Delaware): 2915 Ogletown Road, Newark De 19713, EE.UU. Fort-Lauderdale (Florida): 3100 Ray Ferrero Jr. Blvd, Fort Lauderdale, FL 33314 , Sala 5043, EE.UU.
Duración del proyecto	2016- actualidad
DETALLES TÉCNICOS	
Tipo de IA Utilizada	Aprendizaje profundo (deep learning) y aprendizaje automático (machine learning)
Algoritmos o modelos	Algoritmos de clasificación supervisada: Utilizan técnicas de aprendizaje automático para etiquetar automáticamente puntos de una nube. Redes neuronales: Estas redes pueden utilizarse para entrenar modelos de clasificación basados en características geométricas y de color de los puntos LiDAR. Algoritmos basados en reglas: Estos algoritmos aplican reglas predefinidas (como la altura, la pendiente o el color) para clasificar los puntos.
Datos utilizados	Datos LIDAR y archivos de nubes de puntos (.las, .laz, .e57)

DESCRIPCIÓN DE LOS RESULTADOS OBTENIDOS	
Los resultados que ofrece el software permiten no solo la visualización detallada del entorno, sino también el análisis preciso de elementos específicos como volúmenes, estructuras, terrenos y vegetación. Estas capacidades hacen que el software sea una herramienta esencial para diversos profesionales que trabajan con datos LIDAR. Se han obtenido modelos 3D, que gracias al uso de la IA se han comparado rápidamente, obteniendo los resultados prácticamente en tiempo real.	
MEJORAS EN LA EMPRESA O INSTITUCIÓN	
Permitió reducir el tiempo de obtención de datos, ya que antiguamente en este proyecto se realizaba con estación total y analizando manualmente. La diferencia es demasiada en prácticamente todos los aspectos. Haciendo que esta nueva metodología sea más precisa y eficiente.	
INDICADORES DE ÉXITO	
• Rapidez en la obtención y análisis • Análisis más exactos y con más datos • Mejora en la seguridad y salud de la compañía	
DESAFÍOS Y SOLUCIONES	
Problemas enfrentados	Los desafíos incluyen: • Densidad variable: Las nubes de puntos pueden tener diferentes densidades en distintas áreas. • Ruido: Puntos no deseados pueden interferir con la clasificación. • Variabilidad: Diferentes formas y tamaños de objetos pueden dificultar la clasificación. • Inversiones iniciales significativas
Cómo se superaron	Filtrado de ruido: Este algoritmo elimina puntos de la nube que se consideran "ruido", es decir, puntos que no pertenecen a las estructuras reales capturadas. Downsampling: Reduce la cantidad de puntos en la nube para hacer más manejables los datos, manteniendo la precisión en las áreas críticas. La inversión es rentable ya que a largo plazo ya que un proceso manual con costes se vuelve automático sin apenas tenerlos.
LECCIONES APRENDIDAS	
Aspectos positivos	La calidad y representatividad del dataset de entrenamiento es crucial para el éxito del modelo de IA y que los resultados obtenidos sean de mayor fiabilidad para poder analizarlos y hallar conclusiones en base a estos. La tecnología LIDAR es capaz de ayudar a sectores como el de la minería a mejorar la calidad de trabajo y permitir que este crezca junto a las nuevas tecnologías y la IA.

Áreas de mejora	• Mayor frecuencia en la toma de datos para llevar un seguimiento más exhaustivo de las deformaciones en la mina. • Mejora en los resultados que puede otorgar los algoritmos de la IA. • Se podría entrenar a la IA para que decida dónde colocar los barrenos dependiendo de la estructura interna de la mina.
PROYECCIONES FUTURAS	
Planes de expansión o mejora	Colocar y programar un dispositivo que obtenga los datos LIDAR en una vagoneta o similar en una mina subterránea, o, un dron en una mina abierta que obtenga los datos y los procese de forma automática para obtener los movimientos con una periodicidad mayor.
Posibles nuevas aplicaciones	Implementación de un dispositivo permanente que recoja datos para trabajos de auscultación

CONCLUSIÓN	
El LIDAR está sentando las bases para una nueva era en la minería, caracterizada por operaciones más seguras, eficientes y sostenibles. A medida que la industria continúa evolucionando, la adopción de tecnologías avanzadas como LIDAR será fundamental para mantener la competitividad y cumplir con los estándares ambientales y de seguridad.	
REFERENCIAS Y RECURSOS ADICIONALES	
Enlaces a documentación y estudios de caso	https://geo-plus.com/es/software-nube-de-puntos/visionlidar-estudios-de-caso/ 
Contacto para más información	https://www.visionlidar.co/contact-g
FOTOGRAFÍA	

ONBOARDING DIGITAL

 José Manuel Linares

TITULO DEL CASO DE ÉXITO	
Proyecto onboarding digital. Investigación en técnicas de Machine Learning como aprendizaje Few-Shot, Transferencia de estilo neuronal y Super Resolución Multi-Frame para mejorar rendimiento de modelos de PAD (Presentation Attack Detection).	
DESCRIPCIÓN BREVE DEL PROYECTO	
Resumen del problema o necesidad	El aumento de fraudes bancarios como phishing, spoofing y ataques de presentación (PA) ha generado la necesidad de herramientas avanzadas para proteger a las entidades financieras y a los usuarios. Los delincuentes utilizan tecnologías para falsificar documentos y suplantar identidades, lo que afecta la confianza y seguridad en los sistemas de pagos digitales. Con la implementación de la normativa DORA por la Unión Europea, las instituciones deben garantizar mayor resiliencia digital y ciberseguridad.
Solución implementada con IA	Facephi e ITI desarrollaron una solución basada en IA generativa y Deep Learning para combatir los ataques de presentación mediante onboarding digital. La tecnología utiliza modelos de IA y Neural Style Transfer para crear documentos sintéticos y entrenar sistemas de detección. Estos modelos permiten verificar la autenticidad de documentos de identidad y detectar fraudes de forma automatizada, logrando mejorar la seguridad en la adquisición de clientes de forma remota.
CONTEXTO DEL PROYECTO	
Industria o sector	Sector financiero y tecnología digital
Localización	Alicante
Duración del proyecto	12 meses
DETALLES TÉCNICOS	
Tipo de IA Utilizada	IA generativa, Deep Learning y Neural Style Transfer
Algoritmos o modelos	Se utilizaron cuatro modelos basados en redes generativas adversariales (GAN): pix2pix, pix2pixHD, CycleGAN y CUT.

Herramientas y plataformas	Los modelos se implementaron utilizando el lenguaje de programación python3, basándose en la librería Pytorch Lightning y se ejecutaron en un servidor dedicado para la investigación en Machine Learning con 32 núcleos, 236 GB de RAM y 40 GB de GPU.
Datos utilizados	Se utilizaron los datasets MIDV-2020 y DLC-2021, que contienen clips de video de documentos de identidad falsos en diversas condiciones de iluminación y fondos. Se utilizaron aproximadamente 48,350 frames de estos vídeos y se generaron 21,700 imágenes sintéticas mediante modelos GAN.

DESCRIPCIÓN DE LOS RESULTADOS OBTENIDOS	
Prototipos de modelos generativos para el aumentado de datos de documentos de identificación. Las muestras generadas mejoran el rendimiento de los modelos de PAD, pero el mejor método generativo depende del tipo de ataque. Mejora en la seguridad de transacciones digitales y confianza de los usuarios, optimización del proceso de onboarding y cumplimiento regulatorio de ciberseguridad.	
MEJORAS EN LA EMPRESA O INSTITUCIÓN	
Este proyecto aporta a Facephi mejoras en la detección de fraudes, ya que la creación de documentos sintéticos permite entrenar sistemas más eficientes contra ataques de presentación (PA). Además, optimiza el proceso de onboarding digital al automatizar la verificación de identidad, lo que mejora la experiencia del usuario.	
INDICADORES DE ÉXITO	
(Estadísticas, métricas de rendimiento, ROI, etc.)	
DESAFÍOS Y SOLUCIONES	
Problemas enfrentados	Los principales desafíos incluyeron la dificultad para obtener conjuntos de datos representativos debido a restricciones legales y la naturaleza privada de los documentos de identificación. También la generación de imágenes sintéticas, donde problemas de alineación y calidad visual afectaron los resultados. Del mismo modo, fue difícil lograr un balance entre la mejora del rendimiento de los modelos de detección de fraudes y la fidelidad de las imágenes generadas, especialmente cuando se usaban datos sintéticos en lugar de reales.

Cómo se superaron	Se generaron imágenes sintéticas que complementarían los datos reales, abordando así la escasez de datos reales de ataques de presentación. Para facilitar el entrenamiento de los modelos supervisados, implementaron un proceso de alineación automática de imágenes mediante el algoritmo ORB, lo que automatizó la creación de pares de datos bona fide y ataques. Finalmente, evaluaron el impacto de los datos sintéticos comparándolos con datos reales, demostrando que, en ciertos casos, los modelos entrenados con imágenes generadas lograron incluso mejorar el rendimiento en la detección de fraudes.
LECCIONES APRENDIDAS	
Aspectos positivos	La colaboración entre empresas tecnológicas, la innovación en el uso de IA Generativa y Deep Learning para detectar fraudes.
Áreas de mejora	La necesidad de mejorar la calidad de los datos utilizados, tanto de las imágenes sintéticas generadas como de los datasets públicos.
PROYECCIONES FUTURAS	
Planes de expansión o mejora	Continuar innovando y perfeccionando los modelos generativos y de IA para adaptarse a nuevas técnicas de fraude y mejorar la precisión en la detección de documentos falsificados.
Posibles nuevas aplicaciones	Podrían integrar la tecnología de detección de documentos falsificados en aplicaciones móviles para realizar verificaciones de identidad de forma automática.

CONCLUSIÓN

El proyecto de Facephi e ITI ha permitido desarrollar prototipos de modelos generativos para el aumento de datos de documentos de identificación, para mejorar la detección de ataques de presentación (PA).

La creación de documentos sintéticos ha permitido entrenar sistemas de detección más precisos, aumentando la seguridad en las transacciones digitales y fortaleciendo la confianza de los usuarios.

Sin embargo, el proyecto también enfrentó desafíos, como la dificultad en la obtención de datos representativos y la mejora de la calidad visual de las imágenes generadas, que se superaron con innovaciones en alineación automática de imágenes y evaluación comparativa.

Las lecciones aprendidas subrayan la importancia de la colaboración tecnológica y la necesidad de continuar perfeccionando la calidad de los datos y los modelos.

REFERENCIAS Y RECURSOS ADICIONALES

Enlaces a documentación y estudios de caso

Paper: <https://arxiv.org/pdf/2312.13993>



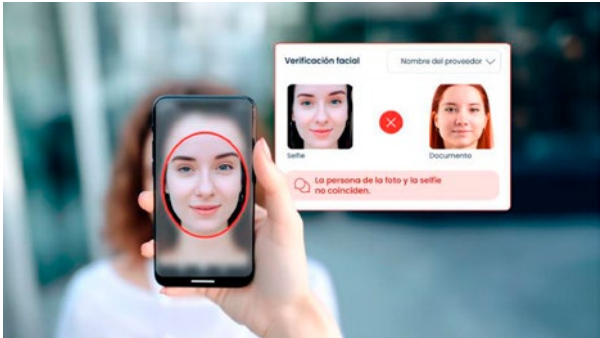
Contacto para más información

Raúl Hussein
rhussein@iti.es

URL / QR

<https://valenciaplaza.com/iti-facephi-desarrollan-herramienta-combatir-fraudes-bancarios-suplantacion-identidades>

Fotografía



GLOSARIO IMPRESCINDIBLE DE LA IA

Conceptos básicos de Inteligencia Artificial

1. **Agente autónomo:** Sistema de IA capaz de operar independientemente, tomar decisiones y realizar acciones sin intervención humana directa.
2. **Agente basado en objetivos:** Agente que toma decisiones basadas en metas específicas.
3. **Agente basado en utilidad:** Agente que toma decisiones para maximizar una función de utilidad.
4. **Agente inteligente:** Sistema autónomo que percibe su entorno y actúa para lograr objetivos.
5. **Agente reactivo:** Agente que responde directamente a los estímulos sin mantener un estado interno.
6. **Agente reflexivo:** Agente que actúa según reglas predefinidas basadas en la percepción actual.
7. **AGI:** Acrónimo de Artificial General Intelligence. Véase Inteligencia Artificial General
8. **Algoritmo:** Conjunto de reglas o instrucciones definidas para resolver un problema o realizar una tarea.
9. **Algoritmo genético:** Técnica de optimización inspirada en la evolución biológica.
10. **ANI:** Acrónimo de Artificial Narrow Intelligence, Véase IA débil
11. **Análisis de sentimientos:** Uso de NLP para identificar y extraer opiniones de un texto.
12. **ANN:** Acrónimo de Artificial Neural Networks. Véase Redes Neuronales Artificiales
13. **Aprendizaje automático (Machine Learning):** Subcampo de la IA que permite a las máquinas aprender de los datos sin ser programadas explícitamente.
14. **Aprendizaje continuo:** Capacidad de un sistema de IA para seguir aprendiendo y actualizándose después de su despliegue inicial.
15. **Aprendizaje federado:** Técnica que entrena algoritmos en dispositivos descentralizados.
16. **Aprendizaje no supervisado:** Método de machine learning que trabaja con datos no etiquetados.
17. **Aprendizaje por refuerzo:** Técnica de machine learning donde un agente aprende a tomar decisiones mediante recompensas o penalizaciones.
18. **Aprendizaje profundo (Deep Learning):** Subset del machine learning basado en redes neuronales artificiales con múltiples capas.

19. **Aprendizaje semi-supervisado:** Combina datos etiquetados y no etiquetados en el entrenamiento.
20. **Aprendizaje supervisado:** Tipo de machine learning donde el modelo se entrena con datos etiquetados.
21. **Árbol de decisión:** Modelo de predicción que representa decisiones y sus posibles consecuencias.
22. **Arquitectura de agentes:** Diseño y estructura de un sistema de agentes inteligentes.
23. **Arquitectura de red:** Estructura y diseño de una red neuronal artificial.
24. **Autocodificador:** Red neuronal utilizada para aprendizaje de características y reducción de dimensionalidad.
25. **Automatización cognitiva:** Uso de IA para automatizar tareas que tradicionalmente requerían inteligencia humana.
26. **Automatización robótica de procesos (RPA):** Tecnología que automatiza tareas repetitivas basadas en reglas.
27. **Bayesiano, enfoque:** Método estadístico basado en el teorema de Bayes.
28. **BERT (Bidirectional Encoder Representations from Transformers):** Modelo de lenguaje preentrenado para NLP.
29. **Big Data:** Conjuntos de datos extremadamente grandes y complejos.
30. **Biometría:** Uso de características físicas o conductuales para la identificación.
31. **Chatbot:** Programa de IA diseñado para simular conversaciones con usuarios humanos.
32. **Clasificación:** Tarea de machine learning que asigna datos de entrada a categorías predefinidas.
33. **Clustering:** Técnica de aprendizaje no supervisado que agrupa datos similares.
34. **CNN (Redes Neuronales Convolucionales):** Tipo de red neuronal especialmente eficaz en el procesamiento de imágenes.
35. **Computación afectiva:** Área de la IA que se ocupa del reconocimiento y simulación de emociones humanas.
36. **Computación cognitiva:** Sistemas que simulan procesos de pensamiento humano.
37. **Computación cuántica:** Uso de fenómenos cuánticos para realizar cálculos.
38. **Conjunto de datos:** Colección de datos utilizada para entrenar y evaluar modelos.
39. **Conciencia artificial:** Concepto hipotético de máquinas con autoconciencia similar a la humana.
40. **Cooperación entre agentes:** Interacción y colaboración entre múltiples agentes inteligentes.
41. **Capa oculta:** Capa intermedia en una red neuronal entre la entrada y la salida.
42. **Cross-entropy:** Función de pérdida utilizada en problemas de clasificación.

- 43. **Datos de entrenamiento:** Conjunto de datos utilizado para entrenar un modelo de machine learning.
- 44. **Datos de prueba:** Conjunto de datos utilizado para evaluar el rendimiento de un modelo entrenado.
- 45. **Datos de validación:** Conjunto de datos usado para ajustar hiperparámetros y evaluar el modelo durante el entrenamiento.
- 46. **Deep Learning:** Véase Aprendizaje Profundo
- 47. **Descenso de gradiente:** Algoritmo de optimización utilizado en machine learning.
- 48. **Detección de anomalías:** Identificación de patrones inusuales en los datos.
- 49. **Dropout:** Técnica de regularización para prevenir el sobreajuste en redes neuronales.
- 50. **Embeddings:** Representaciones de baja dimensión de datos de alta dimensión.
- 51. **Entorno de agente:** Contexto en el que opera un agente inteligente.
- 52. **Entropía:** Medida de la incertidumbre o aleatoriedad en los datos.
- 53. **Épocas:** Número de veces que el algoritmo de aprendizaje trabaja a través de todo el conjunto de datos de entrenamiento.
- 54. **Estado:** Descripción completa de la situación actual de un sistema o agente en un momento dado. Representa toda la información relevante que el sistema o agente necesita para tomar decisiones o realizar acciones.
- 55. **Estado interno:** Conjunto de variables o información que un agente o sistema mantiene internamente para representar su situación actual y su historia. Permite a los agentes que no son puramente reactivos, tomar decisiones basadas no solo en la percepción actual, sino también en experiencias pasadas o conocimientos acumulados.
- 56. **Ética de la IA:** Consideraciones morales y éticas en el desarrollo y uso de la IA.
- 57. **Explicabilidad de la IA:** Capacidad de comprender y explicar cómo un sistema de IA llega a sus conclusiones.
- 58. **Feature engineering:** Proceso de selección y transformación de variables para mejorar el rendimiento del modelo.
- 59. **Función de activación:** Función que determina la salida de un nodo en una red neuronal.
- 60. **Función de pérdida:** Mide la diferencia entre la predicción del modelo y el valor real.
- 61. **GAN (Redes Generativas Adversarias):** Arquitectura de IA que genera datos nuevos.
- 62. **Generalización:** Capacidad de un modelo para funcionar bien con datos nuevos y no vistos.
- 63. **GPT (Transformador Generativo Pre-entrenado):** Modelo de lenguaje basado en la arquitectura del transformador.

64. **Gradient boosting:** Técnica de machine learning que combina modelos débiles en un modelo fuerte.
65. **Hiperparámetros:** Parámetros que se establecen antes del proceso de entrenamiento del modelo.
66. **IA débil:** IA diseñada para realizar tareas específicas.
67. **IA fuerte:** Hipotética IA con capacidades cognitivas similares o superiores a las humanas.
68. **IA Generativa:** Sistemas de IA capaces de crear contenido original.
69. **Industria 1.0:** Revolución industrial basada en la mecanización impulsada por energía hidráulica y de vapor, transformando la producción artesanal en manufactura masiva.
70. **Industria 2.0:** Segunda revolución industrial caracterizada por la electrificación, producción en masa mediante líneas de ensamblaje y la expansión del transporte y las comunicaciones.
71. **Industria 3.0:** Etapa marcada por la automatización, electrónica e informática, introduciendo sistemas computarizados y robótica en los procesos industriales para optimizar eficiencia y calidad.
72. **Industria 4.0:** Concepto que describe la transformación industrial impulsada por tecnologías como IoT, IA, Big Data y robótica, promoviendo fábricas inteligentes y automatización avanzada.
73. **Industria 5.0:** Enfoque industrial centrado en la colaboración entre humanos y máquinas, buscando personalización, sostenibilidad y un impacto positivo en la sociedad más allá de la productividad.
74. **IA Simbólica:** Enfoque de IA basado en la manipulación de símbolos y reglas lógicas.
75. **Inferencia:** Proceso de hacer predicciones utilizando un modelo entrenado.
76. **Interfaz cerebro-computadora:** Tecnología que permite la comunicación directa entre el cerebro y dispositivos externos.
77. **IoT (Internet de las Cosas):** Red de dispositivos físicos interconectados que recopilan y comparten datos.
78. **K-means:** Algoritmo de clustering popular en aprendizaje no supervisado.
79. **K-nearest neighbors:** Algoritmo de clasificación y regresión basado en la similitud de características.
80. **LAM (Local Action Model):** Modelo de acción local que mejora la eficiencia operativa al descentralizar la toma de decisiones y permitir acciones rápidas y adaptadas a las necesidades específicas de cada contexto local.
81. **LSTM (Long Short-Term Memory):** Tipo de red neuronal recurrente capaz de aprender dependencias a largo plazo.

- 82 . **Machine Learning:** Véase Aprendizaje Automático
- 83 . **Máquina de vectores de soporte (SVM):** Algoritmo de aprendizaje supervisado usado para clasificación y regresión.
- 84 . **Matriz de confusión:** Herramienta para evaluar el rendimiento de un modelo de clasificación.
- 85 . **Minería de datos:** Proceso de descubrir patrones en grandes conjuntos de datos.
- 86 . **Modelo:** Representación matemática de un sistema o proceso del mundo real.
- 87 . **Módulo de atención:** Componente de redes neuronales que permite enfocarse en partes específicas de la entrada.
- 88 . **Multi-agente, sistema:** Conjunto de agentes inteligentes que interactúan en un entorno compartido.
- 89 . **Naive Bayes:** Algoritmo de clasificación basado en el teorema de Bayes.
- 90 . **NLP (Procesamiento del Lenguaje Natural):** Rama de la IA que se ocupa de la interacción entre computadoras y lenguaje humano.
- 91 . **Neurona artificial:** Unidad básica de procesamiento en una red neuronal artificial.
- 92 . **Normalización:** Proceso de ajustar los valores de las características a una escala común.
- 93 . **One-hot encoding:** Técnica para representar variables categóricas como vectores binarios.
- 94 . **Optimización:** Proceso de ajustar los parámetros de un modelo para mejorar su rendimiento.
- 95 . **Overfitting:** Sobreajuste del modelo a los datos de entrenamiento, perdiendo capacidad de generalización.
- 96 . **Percepción del agente:** Capacidad de un agente para recibir y procesar información de su entorno.
- 97 . **Perceptrón:** Tipo más simple de red neuronal artificial.
- 98 . **PLN (Procesamiento del Lenguaje Natural):** Capacidad de un sistema de IA para comprender, interpretar y generar lenguaje humano.
- 99 . **Planificación de agentes:** Proceso de determinar una secuencia de acciones para alcanzar un objetivo.
- 100 . **Poda (Pruning):** Técnica para reducir el tamaño de los modelos de machine learning.
- 101 . **Política de agente:** Estrategia que un agente utiliza para decidir qué acción tomar.
- 102 . **Primera era de la máquina:** Etapa de la Revolución Industrial caracterizada por la invención de la máquina de vapor y la mecanización de la producción, que transformó la manufactura y la industria en el siglo XIX.
- 103 . **Prueba de Turing:** Test propuesto para evaluar la capacidad de una máquina para exhibir un comportamiento inteligente.

- 104 . **Q-learning:** Algoritmo de aprendizaje por refuerzo para aprender una política de acción óptima.
- 105 . **Random forest:** Algoritmo de ensamble que construye múltiples árboles de decisión.
- 106 . **Razonamiento basado en casos:** Técnica de resolución de problemas que utiliza soluciones pasadas para resolver nuevos problemas.
- 107 . **Reconocimiento de patrones:** Proceso de identificar regularidades y estructuras en los datos.
- 108 . **Red neuronal:** Modelo de machine learning inspirado en la estructura y funcionamiento del cerebro humano.
- 109 . **Regresión:** Tipo de análisis predictivo que estima relaciones entre variables.
- 110 . **Regresión logística:** Modelo estadístico usado para predecir una variable binaria.
- 111 . **Regularización:** Técnica para prevenir el overfitting en modelos de machine learning.
- 112 . **Reinforcement learning:** Véase "Aprendizaje por refuerzo".
- 113 . **Representación del conocimiento:** Forma en que un agente o sistema de IA almacena y organiza la información.
- 114 . **Robótica:** Campo que combina ingeniería y computación para diseñar y construir máquinas automatizadas.
- 115 . **RNA:** Acrónimo de Red Neuronal Artificial. Véase "Red neuronal".
- 116 . **RNN (Redes Neuronales Recurrentes):** Tipo de red neuronal diseñada para trabajar con secuencias de datos.
- 117 . **Segunda era de la máquina:** Período iniciado con la computadora digital, donde las tecnologías de la información y la automatización redefinen la economía y los procesos productivos, aumentando la productividad y creando nuevas formas de trabajo y empleo
- 118 . **Sesgo (Bias):** Error sistemático en un modelo de machine learning.
- 119 . **Singularidad tecnológica:** Hipotético punto futuro en el que la IA supera la inteligencia humana, llevando a cambios impredecibles.
- 120 . **Sistema 1:** En los modelos de lenguaje, se refiere a los modelos que permiten generar predicciones y completar texto de manera fluida y natural, al aprovechar los patrones aprendidos durante el entrenamiento
- 121 . **Sistema 2:** : En los modelos de lenguaje se refiere al proceso de pensamiento más lento, controlado y deliberativo, que se encarga de las tareas más complejas como el análisis semántico profundo, el razonamiento abstracto y la toma de decisiones importantes.
- 122 . **Sistemas expertos:** Programas de IA que emulan el conocimiento y habilidades de un experto humano.

- 123 . **Superinteligencia:** Hipotética IA con capacidades cognitivas muy superiores a las humanas en prácticamente todos los campos.
- 124 . **Supervisión humana:** Participación humana en el desarrollo, entrenamiento y operación de sistemas de IA.
- 125 . **Tecno Paradigma (primero):** Revolución agrícola e hidráulica, marcada por la domesticación de plantas y animales, y la construcción de sistemas de riego para facilitar la agricultura organizada.
- 126 . **Tecno Paradigma (segundo):** Revolución industrial inicial basada en la máquina de vapor, que introdujo la mecanización en sectores como el textil y el transporte, impulsando el desarrollo económico global.
- 127 . **Tecno Paradigma (tercero):** Revolución eléctrica y química, destacada por la electrificación, el desarrollo de productos químicos y la comunicación mediante el telégrafo y el teléfono.
- 128 . **Tecno Paradigma (cuarto):** Era del petróleo y la producción en masa, con avances en la industria automotriz, petroquímica y electrificación masiva.
- 129 . **Tecno Paradigma (quinto):** Revolución digital basada en microprocesadores, informática, telecomunicaciones y tecnologías de la información, transformando economías y sociedades.
- 130 . **Tecno Paradigma (sexto):** En desarrollo, caracteriza una convergencia entre biotecnología, IA, nanotecnología, energías renovables y sostenibilidad, enfocándose en soluciones globales para retos como el cambio climático.
- 131 . **Tensor:** Objeto matemático que describe una relación lineal entre vectores, escalares y otros tensores.
- 132 . **Teoría de juegos:** Estudio de modelos matemáticos de conflicto y cooperación entre agentes racionales.
- 133 . **Transferencia de aprendizaje:** Técnica que permite aplicar el conocimiento adquirido en una tarea a otra tarea relacionada.
- 134 . **Transformers:** Arquitectura de red neuronal basada en mecanismos de atención.
- 135 . **Vector:** Son la forma fundamental de representar y procesar el lenguaje en sistemas de IA modernos.
- 136 . **Visión por computadora:** Campo de la IA que entrena a las computadoras para interpretar y entender el mundo visual.
- 137 . **XAI (IA Explicable):** Enfoque en el desarrollo de sistemas de IA cuyas decisiones pueden ser entendidas y rastreadas por los humanos.
- 138 . **Zona de confort del agente:** Rango de situaciones o estados en los que un agente de IA puede operar de manera efectiva y confiable, basándose en su entrenamiento y diseño.

LISTA DE AUTORES DE LA GUÍA BÁSICA DE LA IA

1. Aguilar Porras, Ivana del Rocío



Ingeniera Técnica de Telecomunicaciones, especialidad Sonido e Imagen.
Ingeniera TIC en Dextromédica S.L.
Colegio Oficial Ingenieros Técnicos de Telecomunicación.
[linkedin.com/in/ivana-aguilar](https://www.linkedin.com/in/ivana-aguilar)

2. Alepuz Moner, Vicente



Graduado en Ingeniería de Sistemas de Telecomunicación,
Sonido e Imagen.
Responsable de I+D+i en Ionclinics.
Colegio Oficial Ingenieros Técnicos de Telecomunicación.
[linkedin.com/in/vicente-alepuz-471520b6](https://www.linkedin.com/in/vicente-alepuz-471520b6)

3. Alfaro Villa, Jesús



Ingeniero de Caminos, Canales y Puertos.
Delegado de firmas Grupo Bertolín y Profesor Asociado UPV.
Colegio Oficial Ingenieros de Caminos, Canales y Puertos.
[linkedin.com/in/jesús-alfaro-villa-988330157](https://www.linkedin.com/in/jesús-alfaro-villa-988330157)

4. Bañón Briet, Jorge Jaime



Ingeniero en Geomática y Topografía.
GlobeArea Control S.L.
Colegio Oficial de Ingenieros de Geomática y Topografía.
[linkedin.com/in/jorge-jaime-bañón-briet-868a131ba](https://www.linkedin.com/in/jorge-jaime-bañón-briet-868a131ba)

5. Botti Navarro, Vicent J.



Ingeniero Informático
Catedrático en la UPV. Director de VRAIN y en ValgrAI
Colegio Oficial de Ingenieros Informáticos
[linkedin.com/in/vicente-botti-82614a17](https://www.linkedin.com/in/vicente-botti-82614a17)

6. Despujol Zabala, Ignacio



Ingeniero de Telecomunicación

Jefe de proyecto nuevas aplicaciones de Internet UPV

Colegio Oficial de Ingenieros de Telecomunicación

[linkedin.com/in/ndespujol](https://www.linkedin.com/in/ndespujol)

7. Giménez Millán, Julio



Economista

Consultor y formador en transformación digital de centros educativos.

Scooltic

Colegio Oficial de Economistas de Valencia (COEV)

[linkedin.com/in/juliogimenezmillan](https://www.linkedin.com/in/juliogimenezmillan)

8. Linares Pellicer, Jordi



Ingeniero Informático

Profesor en UPV. Miembro de VRAIN. Coordinador en ValgrAI

Colegio Oficial de Ingenieros Informáticos

[linkedin.com/in/jlinares](https://www.linkedin.com/in/jlinares)

9. José Manuel Linares Suárez



Ingeniero Industrial

Business Development Manager en el ITI

Colegio Oficial de Ingenieros Informáticos

[linkedin.com/in/josemanuellinaressuarez](https://www.linkedin.com/in/josemanuellinaressuarez)

10. Majadas Morales, Victoria



Ingeniera de Telecomunicaciones

Fundadora en Smart To People. Senior Advisor Business Angel.

Presidenta BIGBAN

Colegio Oficial de Ingenieros de Telecomunicación

[linkedin.com/in/victoriamajadasmorales](https://www.linkedin.com/in/victoriamajadasmorales)

11. Martínez Martínez, Juan



Economista

Consultor y formador en transformación digital e inteligencia artificial

Colegio Oficial de Economistas de Valencia (COEV)

[linkedin.com/in/experinter](https://www.linkedin.com/in/experinter)

12. **Monsoriu Flor, Mar**

Periodista



Escritora y profesora especializada en tecnología.

Experta en Inteligencia Artificial.

Asociación Profesional de Periodistas Valencianos

[linkedin.com/in/monsoriu](https://www.linkedin.com/in/monsoriu)

13. **Montesa Andrés, Enric**

Economista



Consultor en Estudios de Mercado y Opinión

Colegio Oficial de Economistas de Valencia (COEV)

[linkedin.com/in/enricmontesa](https://www.linkedin.com/in/enricmontesa)

14. **Morillas Barrio, César**

Ingeniero de Telecomunicación



Socio fundador Telytec

Colegio de Ingenieros de Telecomunicación

[linkedin.com/mwlite/profile/in/cesar-morillas-barrio](https://www.linkedin.com/mwlite/profile/in/cesar-morillas-barrio)

15. **Muñoz Vela, José Manuel**

Abogado



Director Jurídico Adequa Corporación

Abogado especialista en Derecho Digital e IA. Doctor en Derecho (IA)

[linkedin.com/in/jose-manuel-munoz-vela-phd-4524b27](https://www.linkedin.com/in/jose-manuel-munoz-vela-phd-4524b27)

16. **Ortega Bonilla, José**

Economista



Consultor de dirección, financiero y novelista

Colegio Oficial de Economistas de Valencia (COEV)

[linkedin.com/in/jose-ortega-bonilla-04862912](https://www.linkedin.com/in/jose-ortega-bonilla-04862912)

17. **Ortuño Padilla, Armando**

Ingeniero de Caminos, Canales y Puertos. Economista



Profesor de Territorio del Dpto. de Ingeniería Civil de la Universidad de Alicante

Colegio de Ingenieros de Caminos, Canales y Puertos

[linkedin.com/in/armando-ortuno-padilla-60493b1b](https://www.linkedin.com/in/armando-ortuno-padilla-60493b1b)

18. Peñarrubia Carrión, Juan Pablo



Ingeniero en informática

Investigador en ética e impacto de las tecnologías informáticas

Colegio Oficial de Ingeniería Informática

[linkedin.com/in/jppenarrubia](https://www.linkedin.com/in/jppenarrubia)

19. Plasencia Diago, Adolfo



Periodista

Escritor, profesor y periodista de ciencia y tecnología.

Autor de "De neuronas a galaxias"

[linkedin.com/in/adolfoplasencia](https://www.linkedin.com/in/adolfoplasencia)

20. Rieta Carbonell, Joaquín



Economista

Consultor y formador en innovación con transformación digital e inteligencia artificial.

Colegio Oficial de Economistas de Valencia (COEV)

[linkedin.com/in/chimorieta/](https://www.linkedin.com/in/chimorieta/)

21. Sales Tarancón, Manuel



Ingeniero de Telecomunicación

Consultor de Financial Services IT en Europe, America y Asia

Asociación Valenciana del Bitcoin (AvalBit)

[linkedin.com/in/manuelsales](https://www.linkedin.com/in/manuelsales)

22. Segarra Clara, Ramón



Ingeniero de Telecomunicación

Gerente de Desarrollo de Negocios CCTV en VISIOTECH.

Director de Seguridad

Colegio de Ingenieros de Telecomunicación

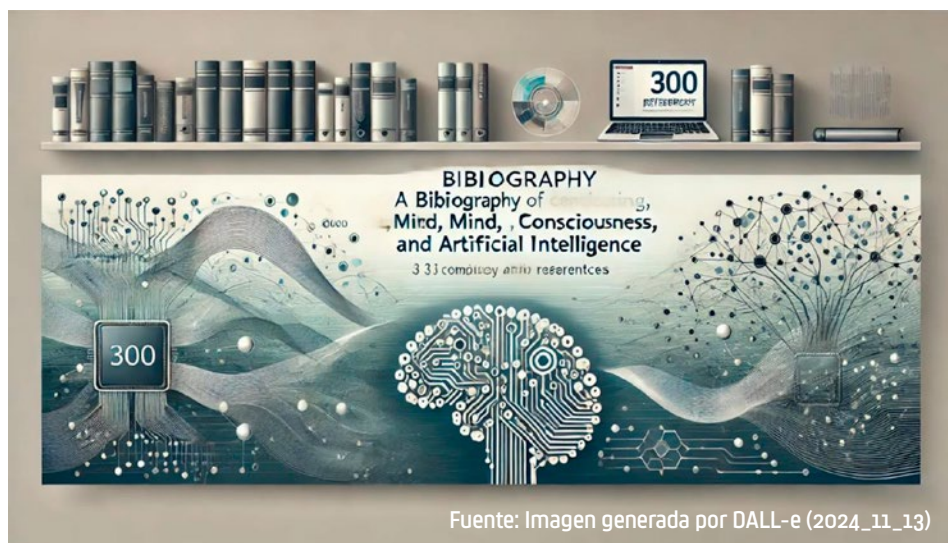
[linkedin.com/in/ramonsegarra](https://www.linkedin.com/in/ramonsegarra)

LISTADO ALFABÉTICO DE LOS COLEGIOS PROFESIONALES QUE INTEGRAN SMART DIGITAL

COLEGIOS SMARTDIGITAL		DIRECCIÓN Y CONTACTO
CICCP	Colegio de Ingenieros de Caminos, Canales y Puertos, Demarcación de la C.V.	C/. Luis Vives, 3 46003 - Valencia Telf.: 96 352 69 61 e-mail: valencia@ciccp.es URL: https://caminoscv.es/
CITOP-CV	Colegio de Ingenieros Técnicos de Obras Públicas (Zona de Valencia y Castellón)	C/ Llosa de Ranes 4, bajo 46021 - Valencia Telf.: 96 339 21 54 e-mail: valencia@citop.es URL: https://www.citopcv.com/
COEVA	Consejo de Colegios de Economistas de la Comunidad Valenciana	C/. Taquígrafo Martí, 4-2 46005 - Valencia Telf.: 96 352 98 69 e-mail: soporte@economistascv.com URL: http://www.economistascv.com/
COGITCV	Colegio Oficial de Graduados e Ingenieros Técnicos de Telecomunicaciones de la CV	C/ Santa Amalia, nº 2 Entlo. 2º; F 46009 - Valencia Telf.: 96 353 10 15 e-mail: cogitcv@cogitcv.org URL: https://itelecoc.es/
COGITI	Colegio Oficial de Ingenieros Técnicos Industriales de Valencia	C/ Guillém de Castro, g. 4º 46007 - Valencia Telf.: 96 351 33 69 e-mail: secretaria@cogitival.es URL: https://www.cogitival.es/
COIGT	Colegio Oficial de Ingeniería Geomática y Topográfica Comunidad Valenciana y Región de Murcia	C/ Dr. Gil y Morte, 25 - Esc. B (Izq.) 1º 46007 - Valencia Telf.: 960 65 97 53 e-mail: cvym@geomaticaytopografia.net URL: https://www.coigt.com/CVYM

COIICV	Colegio Oficial de Ingeniería Informática de la Comunitat Valenciana	C/ Almirante Cadarso 26 46005 - Valencia Telf.: 963 62 29 94 e-mail: secretaria@coiicv.org URL: https://www.coiicv.org/
COIML	Colegio Oficial de Ingenieros de Minas de Levante	C/ Periodista Badía, 6 3º-26 46010 - Valencia Telf.: 96 369 53 49 e-mail: minasleva@gmail.com URL: https://ingenierosdeminasdelevante.org/
COITAVC	Colegio Oficial de Ingenieros Técnicos Agrícolas y Graduados de Valencia y Castellón	C/ Santa Amalia, 2 - Entlo. 1º (Edificio Torres del Turia) 46009 - Valencia Telf.: 963 61 10 15 e-mail: delegacion@coitavc.org
COITCV	Colegio Oficial de Ingenieros de Telecomunicación de la CV	Av. Jacinto Benavente, 12-1ºB 46005 - Valencia Telf.: 96 350 94 94 e-mail: coitcv@coitcv.org URL: https://coitcv.org/
COITICV	Colegio Oficial de Ingenieros Técnicos en Informática de la Comunitat Valenciana	Av. Maisonnave 28 bis 4ª 03003 - Alicante Telf.: 963 62 29 94 e-mail: administracion@coitcv.org URL: https://coitcv.org/
IICV	Colegio Oficial de Ingenieros Industriales de la Comunitat Valenciana	Avda. Francia 55 46023 - Valencia Telf.: 96 351 68 35 e-mail: colegio@iicv.net URL: https://iicv.net/

BIBLIOGRAFÍA



1. Aamodt, S., & Wang, S. (2008): *Entra en tu cerebro*. Barcelona: Ediciones B
2. Acemoglu, D., & Johnson, S. (2023): *Poder y progreso. Nuestra lucha milenaria por la tecnología y la prosperidad*. Barcelona: Ediciones Deusto
3. AFI: Analistas Financieros Internacionales (2024): *Impacto de la IA en la economía española. Ocupaciones, sectores y CC.AA*. AFI. Disponible en: <https://www.afi.es/2024/05/20/inteligencia-artificial-economia-sectores-ocupaciones-comunidades-autonomas/>
4. Agrawal, A., Gans, J., & Goldfarb, A. (2019). *Máquinas predictivas: La sencilla economía de la inteligencia artificial*. Barcelona: Editorial Reverté
5. AI-HLEG (2019): *Ethics guidelines for trustworthy AI. High-Level Expert Group on Artificial Intelligence AI-HLEG*. Versión en español: Directrices éticas para una IA fiable. Bruselas: Comisión Europea, Oficina de Publicaciones. ISBN 978-92-76-11994-4. doi:10.2759/14078. Disponible en: <https://data.europa.eu/doi/10.2759/346720>

6. AI-HLEG (2019): *Ethics guidelines for trustworthy AI. Versión en español: Directrices éticas para una IA fiable*. Bruselas: Comisión Europea, Oficina de Publicaciones
7. Ais Ramírez, D., Mora Illán, T., Pedreño Muñoz, A., Rodríguez Galán, N., & Torres Penalva, A. (2024). *Impulsando el crecimiento de España: el papel de las tecnologías, la IA y las multinacionales*. Alicante: Observatorio de Inteligencia Artificial, Torre Juana OST. Recuperado de <https://ost.torrejuana.es/informe-impulsando-el-crecimiento-de-espana-como-las-tecnologias-la-ia-y-las-multinacionales-estan-transformando-la-economia/>
8. Allison, G. (2018): *Destined For War: can America and China escape Thucydides' Trap?*. Mariner
9. Amooore, L. (2020): *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. Durham (Carolina del Norte): Duke University Press
10. Amooore, L., & Piotukh, V. (2016): *Algorithmic Life: Calculative devices in the age of big data*. New York: Routledge
11. AMTEGA: Axencia para a Modernización Tecnolóxica de Galicia (2023): *Guía práctica para la gestión de la inteligencia artificial en las administraciones públicas*. Santiago de Compostela: Xunta de Galicia. Disponible en: <https://amtega.xunta.gal/es/guia-practica-para-la-gestion-de-la-ia-en-las-AAPP>
12. Ananthaswamy, A. (2024). *Why machines learn: The elegant math behind modern AI*. Barcelona: Penguin Random House.
13. Anderson, Alan (Ed.). (1984). *Controversia sobre mentes y máquinas*. Barcelona: Tusquets Editores
14. Aunger, R. (2004): *El meme eléctrico: Una nueva teoría sobre cómo pensamos*. Barcelona: Paidós
15. Banco Mundial (2017): *Evolución del PIB la PPA mundial en dólares (a precios internacionales constantes de 2017)* <https://datos.bancomundial.org/indicador/NY.GDP.MKTP.PP.KD?end=2023>
16. Banco Mundial (2023): *Evolución del PIB la PPA mundial en dólares (a precios internacionales corrientes de 2023)* <https://datos.bancomundial.org/indicador/NY.GDP.MKTP.CD?end=2023>

17. Baños, G. (2024): *El sueño de la Inteligencia Artificial. El proyecto de construir máquinas pensantes: una historia de la IA*. Barcelona: Shackleton Books
18. Barrio, M. (2023): *Novedades en la tramitación del próximo Reglamento europeo de inteligencia artificial*. Madrid: Real Instituto Elcano
19. Bauer, A., & Raufer, X. (2020): *Geopolítica de la inteligencia artificial: Cómo la revolución digital cambiará nuestras vidas*. Madrid: Ediciones Akal
20. Bear, M. F., Connors, B. W., & Paradiso, M. A. (2015): *Neuroscience: Exploring the Brain*. Lippincott Williams & Wilkins
21. Beckermann, A., McLaughlin, B. P., & Walter, S. (2009): *The Oxford Handbook of Philosophy of Mind*. Oxford (UK): Oxford University Press
22. Benjamins, R., & Salazar, I. (2020): *El mito del algoritmo: Cuentos y cuentas de la inteligencia artificial*. Madrid: Anaya Multimedia
23. Berzal, F. (2018): *Redes Neuronales & Deep Learning*. Edición independiente
24. Boden, M. A. (2017): *Inteligencia Artificial*. Madrid: Turner
25. Bonete Vizcaíno, F. (2024): *La guerra imaginaria*. Barcelona: Siglo XXI editores
26. Bostrom, N. (2016): *Superinteligencia: Caminos, peligros, estrategias*. Madrid: Teell Editorial
27. Bradford, A. (2020): *The Brussels Effect: How the European Union Rules the World*. Oxford University Press
28. Braidot, N. (2015): *Cómo funciona tu cerebro para Dummies*. Barcelona: Grupo Planeta
29. Brundage, M., et al. (2018): *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*. arXiv preprint arXiv:1802.07228. <https://arxiv.org/pdf/1802.07228.pdf>
30. Brynjolfsson, E., & McAfee, A. (2011): *Race Against the Machine*. Digital Frontier Press
31. Brynjolfsson, E., & McAfee, A. (2016): *La segunda era de las máquinas: Trabajo, progreso y prosperidad en una época de brillantes tecnologías*. Buenos Aires: Temas Grupo Editorial

32. Brynjolfsson, E., & McAfee, A. (2017): *Machine, Platform, Crowd: Harnessing Our Digital Future*. W. W. Norton & Company
33. Brynjolfsson, E., & Saunders, A. (2009): *Wired for Innovation*. MIT Press
34. Brzezinski, Z. (1997): *El gran tablero mundial: La supremacía estadounidense y sus imperativos geoestratégicos*. Barcelona: Editorial Paidós
35. Buczak, A. L., & Guven, E. (2016): *A survey of data mining and machine learning methods for cyber security intrusion detection*. IEEE Communications Surveys & Tutorials, 18(2), 1153-1176.
36. Caballero, R., & Martín, E. (2022): *Las bases de Big Data y de la inteligencia artificial*. Madrid: Catarata
37. Cartwright, H. 3ed. (2021): *Artificial Neural Networks*. New York: Humana Press (Springer)
38. Castellón, J. J. (2023): *Confucio: maestro de moral, filósofo de ética*. Sevilla: Isidorianum
39. Castells, M. (1997). *La era de la información: Economía, sociedad y cultura*. Vol. 1, La sociedad red. Madrid: Alianza Editorial
40. Castells, M. (1998). *La era de la información: Economía, sociedad y cultura*. Vol. 2, El poder de la identidad. Madrid: Alianza Editorial
41. Castells, M. (1999). *La era de la información: Economía, sociedad y cultura*. Vol. 3, Fin de milenio. Madrid: Alianza Editorial
42. Castells, M. (2009). *Comunicación y poder*. Madrid: Alianza Editorial
43. Castells, M. (2012). *Redes de indignación y esperanza: Los movimientos sociales en la era de Internet*. Madrid: Alianza Editorial.
44. Chalmers, J. D. (1999): *La mente consciente. En busca de una teoría fundamental*. Barcelona: Gedisa editorial
45. Chalmers, J. D. (2002): *Philosophy of Mind. Classical and Contemporary Readings*. Oxford (UK): Oxford University Press
46. Chalmers, J.D. (2010): *The character of consciousness*. Oxford (UK): Oxford University Press
47. Chalmers, J.D. (2012): *Constructing the World*. Oxford (UK): Oxford University Press

48. Chen, L., & Wang, H. (2022): *Advancements in AI-Powered Video Surveillance*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(8), 1567-1583.
49. Chomón Pérez, J. M. (2023): *La era de las tierras raras: La cruzada geopolítica por los metales estratégicos*. Madrid: Tecnos
50. Clark, A., & Chalmers, D. J. (2011). *La mente extendida*. CIC Cuadernos de Información y Comunicación, 16, 15-28, DOI: https://doi.org/10.5209/rev_CIYC.2011.v16.1
51. CNN (2024): *Ucrania enfrenta problemas con Starlink mientras Rusia aumenta su uso de drones*. <https://cnnespanol.cnn.com/2024/03/26/ucrania-starlink-guerra-drones-rusia-uso-elon-musk-trax/#0>
52. Coeckelbergh, M. (2021): *Ética de la inteligencia artificial*. Madrid: Ediciones Cátedra
53. Coeckelbergh, M. (2023): *La filosofía política de la inteligencia artificial: Una introducción*. Madrid: Ediciones Cátedra
54. Comisión Europea (2020). *Libro Blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza*. COM(2020) 65 final. Bruselas: 19.2.2020. Disponible en: <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020DC0065&from=ES>
55. Comisión Europea (2023): *C(2023)3215 – Standardisation request M/593 commission implementing decision of 22.5.2023 on a standardisation request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in support of Union policy on artificial intelligence*. Comisión Europea. Disponible en https://ec.europa.eu/growth/tools-databases/enorm/mandate/593_en
56. Comisión Europea (2024a): *Public Sector Tech Watch: Mapping Innovation in the EU Public Services*. Bruselas; Publications Office of the European Union. ISBN 978-92-68-11627-2, DOI: 10.2799/4393
57. Comisión Europea (2024b): *Reglamento (UE) 2024/1689 Del Parlamento y del Consejo, de 13 de junio de 2024 por el que se establecen normas armonizadas en materia de inteligencia artificial*. Bruselas: Diario

Oficial de la Unión Europea. Disponible en: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>

- 58 . Comisión Europea (n.d.): *Fondo de Innovación: Licitación competitiva* https://climate.ec.europa.eu/eu-action/eu-funding-climate-action/innovation-fund/competitive-bidding_en?prefLang=es
- 59 . Congreso de los Diputados. (2024). *Inteligencia Artificial. Nota documental Núm. 1*. Madrid: Dirección de Documentación, Biblioteca y Archivo, Departamento de Documentación. Disponible en: https://www.congreso.es/docu/docum/ddocum/notasdocumentales/nd1/inteligencia_artificial.pdf
- 60 . Copeland, J. (1996). *Inteligencia artificial: Una introducción filosófica*. Madrid: Alianza Editorial
- 61 . Copeland, J., Bowen, J., Sprewak, M., & Wilson, Robin (2017): *The Turing Guide*. Oxford: Oxford University Press
- 62 . Coullaut, A., & Tascón, M (2016): *Big Data y el Internet de las cosas*. Madrid: Catarata
- 63 . Council of the European Union (2024): *Artificial Intelligence Act: Council Gives Final Green Light to the First Worldwide Rules on AI* <https://www.consilium.europa.eu/es/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/>
- 64 . Crawford, K. (2022). *Atlas de Inteligencia Artificial. Poder, política y costos planetarios*. 1ed. Ciudad Autónoma de Buenos Aires: Fondo de Cultura Económica
- 65 . Crick, F. (1994): *La búsqueda científica del alma*. Barcelona: Debate
- 66 . Damasio, A. (2001): *La sensación de lo que ocurre*. Barcelona. Destino
- 67 . Damasio, A. (2005): *En busca de Spinoza*. Barcelona: Destino
- 68 . Damasio, A. (2010): *Y el cerebro creó al hombre*. Barcelona: Destino
- 69 . Damasio, A. (2011): *El error de Descartes*. Barcelona: Destino
- 70 . Damasio, A. (2018): *El extraño orden de las cosas*. Barcelona: Destino
- 71 . Damasio, A. (2021): *Sentir y saber*. Barcelona: Destino

72. Davenport, T. H. (2014). *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities*. Harvard Business Review Press.
73. Davenport, T. H. (2018). *The AI Advantage: How to Put the Artificial Intelligence Revolution to Work*. The MIT Press.
74. Davenport, T. H., & Prusak, L. (1999): *Working Knowledge: How Organizations Manage What They Know*. Harvard Business Press
75. Davenport, T. H., Brynjolfsson, E., & McAfee, A. (2019): *Artificial Intelligence: The Insights You Need from Harvard Business Review* (HBR Insights Series). Harvard Business Review Press
76. Davenport, T. H., Iansiti, M., & Neeley, T. (2023): *HBR's 10 Must Reads on AI*. Harvard Business Review Press
77. Dennett, D. C. (1995). *La conciencia explicada*. Barcelona: Ediciones Paidós
78. Dennett, D. C. (1996). *Tipos de mente: Hacia una comprensión de la conciencia*. Madrid: Debate.
79. Dennett, D. C. (2004). *La evolución de la libertad*. Barcelona: Ediciones Paidós
80. Dennett, D. C. (2009). *Romper el hechizo: La religión como un fenómeno natural*. Madrid: Katz Barpal Editores
81. Dennett, D. C. (2017). *De las bacterias a Bach: La evolución de la mente*. Barcelona: Pasado y Presente Ediciones
82. Di Bello, B. (2024): *IA generativa para Dummies*. HOEPLI ediciones
83. Diario Oficial de la Unión Europea (2024): *Reglamento (UE) 2024/1689*. https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=OJ:L_202401689
84. Dilek, S., Çakır, H., & Aydın, M. (2015): *Applications of artificial intelligence techniques to combating cyber crimes: A review*. International Journal of Artificial Intelligence & Applications, 6(1), 21-39.
85. Doidge, N. (2007): *The Brain That Changes Itself: Stories of Personal Triumph from the Frontiers of Brain Science*. Penguin Books

- 86 . Domingo, A., & Peñarrubia, J.P. (2024a): *Herramientas y ética profesional para impulsar la ética de la IA en la empresa*. Proceedings XXXI Congreso Internacional European Business Ethics Network EBEN-Spain 2024
- 87 . Domingo, A., & Peñarrubia, J. P. (2024a): *Herramientas y ética profesional para impulsar la ética de la IA en la empresa*. Proceedings XXXI Congreso Internacional European Business Ethics Network EBEN-Spain 2024: *Business humanism in the digital age: An ethical perspective on artificial intelligence*. 3-5 Junio 2024. Cáceres. Open Access DOI: 10.3926/eben24 / ISBN: 978-84-126475-9-4 Disponible en: <https://www.omniascience.com/books/index.php/proceedings/catalog/book/148>
- 88 . Domingo, A., & Peñarrubia, J. P. (2024b): *Faros en la niebla: Herramientas de ética aplicada para sistemas de IA*. XXV World Congress of Philosophy: "Philosophy across Boundaries". 1-8 Agosto 2024. Roma. <https://wcpro-me2024.com>
- 89 . Domingos, P. (2015): *The Master Algorithm*. Basic Books
- 90 . Dreyfus, H. L. (1965). *Alchemy and Artificial Intelligence* (P-3244). RAND Corporation. Disponible en: <https://www.rand.org/content/dam/rand/pubs/papers/2006/P3244.pdf>
- 91 . Dreyfus, H. L. (1972). *What Computers Can't Do: The Limits of Artificial Intelligence*. Harper & Row
- 92 . Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. Free Press
- 93 . Dreyfus, H. L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*. MIT Press
- 94 . Eagleman, D. (2013): *Incógnito. Las vidas secretas del cerebro*. Barcelona: Anagrama.
- 95 . Eagleman, D. (2017): *El cerebro. Nuestra historia*. Barcelona: Anagrama
- 96 . Eagleman, D. (2024): *Una red viva. La historia interna de nuestro cerebro en cambio permanente*. Barcelona: Anagrama.
- 97 . Engineering News-Record (ENR) (2023): *Top 250 International Contractors Preview* <https://www.enr.com/toplists/2023-Top-250-International-Contractors-Preview>

- 98 . Estulin, D. (2013): *El Club de los Inmortales*. Barcelona: Ediciones B.
- 99 . Estulin, D. (2014): *TransEvolution: The Coming Age of Human Deconstruction*. Trine Day.
- 100 . Estulin, D. (2020): *Metapolítica: La política del futuro*. Barcelona: Ediciones B.
- 101 . Estulin, D. (2022): *El destino de la humanidad*. Barcelona: Libros Cúpula (Planeta)
- 102 . Etxebarria Ecenarro, V. (2022, 7 de septiembre). *Qué es y qué no es inteligencia artificial*. Universidad del País Vasco/Euskal Herriko Unibertsitatea. Cathedra. <https://www.ehu.eus/es/web/campus/home>
- 103 . European Commission (2023): *European Chips Act*. <https://digital-strategy.ec.europa.eu/es/policies/european-chips-act>
- 104 . European Commission (n.d.): *Science Rocks: Rare Earths Technology We Can't Live Without*. <https://projects.research-and-innovation.ec.europa.eu/en/horizon-magazine/science-rocks-rare-earth-technology-cant-live-without>
- 105 . European Union (2020): *Publication on the European Union's Research and Innovation Projects*. <https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>
- 106 . Farnós, J. D. (2024, 8 de julio). *Investigando el «perceptron» (red neuronal artificial) y su desarrollo, dentro del campo automático e inteligente con la Educación disruptiva & IA-AGI*. Juandon. Innovación y conocimiento. <https://juandomingofarnos.wordpress.com/2024/07/08/investigando-el-perceptron-red-neuronal-artificial-y-su-desarrollo-dentro-del-campo-automatico-e-inteligente-con-la-educacion-disruptiva-ia-agi/>
- 107 . Fei Fei Li (2023): *The Worlds I See: Curiosity, Exploration, and Discovery at the Dawn of AI*. Flatiron Books
- 108 . Fernández de Guevara, J., & Minguez, C. (2020): *La inteligencia artificial en la Comunitat Valenciana: medición y propuestas para su impulso*. Valencia: IVIE GVA

- 109 . Flores Galea, A. (2024): *Una mente infinita. La revolución de la inteligencia artificial*. Barcelona: Tusquets editors
- 110 . Floridi, Luciano (2021): *The End of an Era: from Self-Regulation to Hard Law for the Digital Industry*. Philosophy and Technology, 34(4), 619–622. <https://doi.org/10.1007/s13347-021-00493-0>
- 111 . Floris Margadant, G. (2022): *Panorama de la historia universal del derecho*. Barcelona: Marcial Pons
- 112 . Foerster, H. von (1991). *Las semillas de la cibernética*. 1. ed. Barcelona: Gedisa. ISBN: 978-84-7432-414-3
- 113 . Foerster, H. von (2003). *Understanding Understanding: Essays on Cybernetics and Cognition*. New York: Springer-Verlag.
- 114 . García de Cortázar, F. (2023): *Érase una vez Europa: Senderos de justicia, tolerancia y libertad*. Madrid: Espasa
- 115 . Garde, J. L. (2020): *Inteligencia Artificial: Fundamentos y aplicaciones*. Barcelona: Marcombo
- 116 . Gardner, H. (2016): *Estructuras de la mente. La teoría de las inteligencias múltiples*. Ciudad de México: Fondo de Cultura Económica
- 117 . Gawdat, Mo (2021): *La inteligencia que asusta. El futuro de la inteligencia artificial y cómo podemos salvar nuestro mundo*. Barcelona: Paidós.
- 118 . Gobierno de España. (2024). *Estrategia de Inteligencia Artificial 2024*. Madrid. Ministerio para la Transformación Digital y de la Función Pública. https://portal.mineco.gob.es/es-es/digitalizacionIA/Documents/Estrategia_IA_2024.pdf
- 119 . Gobierno de España. (2020). *ENIA: Estrategia Nacional de Inteligencia Artificial*. Madrid: Ministerio de Asuntos Económicos y Transformación Digital. <https://portal.mineco.gob.es/es-es/digitalizacionIA/Paginas/ENIA.aspx>
- 120 . Gobierno de la República Popular China (2017): *Política sobre el desarrollo de la industria de servicios modernos*. https://www.gov.cn/zhen...content_5211996.htm
- 121 . Goodfellow, I., Bengio, Y., & Courville, A. (2016): *Deep Learning*. MIT Press

- 122 . Gumbrecht, H. U. (2020): *El espíritu del mundo en Silicon Valley*. Barcelona: Deusto
- 123 . Harari, Y. N. (2014): *Sapiens: De animales a dioses*. Barcelona: Debate
- 124 . Harari, Y. N. (2016): *Homo Deus: Breve historia del mañana*. Barcelona: Debate
- 125 . Harari, Y. N. (2018): *21 lecciones para el siglo XXI*. Barcelona: Debate
- 126 . Harari, Yuval Noah (2024): *Nexus: Una breve historia de las redes de información desde la Edad de Piedra hasta la IA*. Barcelona: Debate
- 127 . Harvard Business Review, Mollick, E., De Cremer, D., Neeley, T., & Sinha, P. (2024): *Generative AI: The Insights You Need from Harvard Business Review*. Harvard Business Review Press. Recuperado de. <https://store.hbr.org/product/generative-ai-the-insights-you-need-from-harvard-business-review/10697>
- 128 . Hofstadter, D. R. (2015). *Gödel, Escher, Bach: un Eterno y Grácil Bucle*. Barcelona: Tusquets Editores.
- 129 . Hofstadter, D. R., & Dennett, D. C. (1983). *El ojo de la mente: fantasías y reflexiones sobre el yo y el alma*. Buenos Aires: Editorial Sudamericana.
- 130 . Hofstadter, D. R. (2008). *Yo soy un extraño bucle*. Barcelona: Tusquets Editores.
- 131 . Hofstadter, D. R., & Sander, E. (2018). *La analogía: el motor del pensamiento*. Barcelona: Tusquets Editores.
- 132 . ICEX (2025): *Plan 2025: China*. https://www.observatoriorli.com/docs/CHINA/PLAN_2025_China.pdf
- 133 . Imbrie, A., & Kania, E. B. (2021): *El dilema de la inteligencia artificial: Superinteligencia, riesgos nacionales y seguridad internacional*. Barcelona: Ediciones Deusto
- 134 . Immortality Institute. (2004). *The scientific conquest of death: Essays on infinite lifespans* (1a. ed.). Buenos Aires: LibrosEnRed
- 135 . ISO (2022a): *ISO/IEC 22989:2022 Information technology — Artificial intelligence — Artificial intelligence concepts and terminology*. International Organization for Standardization. Disponible en: <https://www.iso.org>

- 136 . ISO (2022b): *ISO/IEC 38507:2022 Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations*. International Organization for Standardization. Disponible en: <https://www.iso.org>
- 137 . ISO (2022c): *ISO/IEC TR 24368:2022 Information technology — Artificial intelligence — Overview of ethical and societal concerns*. International Organization for Standardization. Disponible en: <https://www.iso.org>
- 138 . ISO (2023a): *ISO/IEC 23894:2023 Information technology — Artificial intelligence — Guidance on risk management*. International Organization for Standardization. Disponible en: <https://www.iso.org>
- 139 . ISO (2023b): *ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system*. International Organization for Standardization. Disponible en: <https://www.iso.org>
- 140 . ISO (2023c): *ISO/IEC AWI TS 22443 Information technology — Artificial intelligence — Guidance on addressing societal concerns and ethical considerations*. International Organization for Standardization. Disponible en: <https://www.iso.org>
- 141 . Johnson, S. (2006): *La mente de par en par*. México: Turner - Fondo de Cultura Económica
- 142 . Joyanes Aguilar, L. (2013): *Big Data. Análisis de grandes volúmenes de datos en organizaciones*. Madrid: Alfaomega Grupo Editor
- 143 . Joyanes Aguilar, L. (2017): *Industria 4.0. La cuarta revolución industrial*. Barcelona: Marcombo
- 144 . Joyanes Aguilar, L. (2023). *Ciencias de datos*. Barcelona: Marcombo.
- 145 . Jubak, J. (1993): *La máquina pensante: El cerebro humano y la inteligencia artificial*. Barcelona: Ediciones B
- 146 . Kahneman, D. (2012): *Pensar rápido, pensar despacio*. Barcelona: Debate
- 147 . Kahneman, D., & Tversky, A. (2018): *Elección, valores y marcos*. Barcelona: Marcial Pons
- 148 . Kahneman, D., Sibony, O., & Sunstein, C. R. (2021): *Ruido: Un fallo en el juicio humano*. Barcelona: Debate

- 149 . Kai-Fu Lee & Qiuhan, C. (2021): *AI 2041: Ten Visions for Our Future*. Crown Currency
- 150 . Kai-Fu Lee (2020): *Superpotencias de la inteligencia artificial: China, Silicon Valley y el nuevo orden mundial*. Barcelona: Deusto
- 151 . Kaku, M. (2014). *El futuro de nuestra mente: El reto científico para entender, mejorar y fortalecer nuestra mente*. Barcelona: Editorial Debate
- 152 . Kaku, M. (2018). *El futuro de la humanidad. La colonización de Marte, los viajes interestelares, la inmortalidad y nuestro destino más allá de la Tierra*. Barcelona: Editorial Debate
- 153 . Kaloudi, N., & Li, J. (2020): *The AI-based cyber threat landscape: A survey*. ACM Computing Surveys (CSUR), 53(1), 1–34.
- 154 . Kaplan, J. (2015): *Abstenerse humanos: Guía para la riqueza y el trabajo en la era de la inteligencia artificial*. Madrid: Teell Editorial
- 155 . Kaplan, J. (2016): *Artificial Intelligence: What Everyone Needs to Know*. Oxford University Press
- 156 . Kaplan, J. (2024): *Generative Artificial Intelligence: What Everyone Needs to Know*. Oxford University Press
- 157 . Kaplan, R. D. (2012): *The Revenge of Geography: What the Map Tells Us About Coming Conflicts and the Battle Against Fate*. Random House
- 158 . Khanna, P. (2019): *Geopolítica de la transformación global: Desafíos y oportunidades de un mundo interconectado*. Barcelona: Ediciones Deusto
- 159 . Khanna, P. (2019): *El siglo de Asia: La historia económica del dominio global de Asia*. Barcelona: Editorial Planeta
- 160 . Kissinger, H., Schmidt, E., & Hurrenlocher, D. (2023): *La era de la Inteligencia Artificial y nuestro futuro humano*. Madrid: Anaya Multimedia
- 161 . Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *ImageNet classification with deep convolutional neural networks*. Advances in Neural Information Processing Systems, 25. <https://papers.nips.cc/paper/4824-image-net-classification-with-deep-convolutional-neural-networks.pdf>
- 162 . Kurzweil, R. (1999): *La era de las máquinas espirituales. Cuando los ordenadores superen la mente humana*. Barcelona: Planeta

- 163 . Kurzweil, R. (2012): *La Singularidad está cerca: Cuando los humanos transcendamos la biología*. Berlín: Lola Books
- 164 . Kurzweil, R. (2013): *Cómo crear una mente: El secreto del pensamiento humano*. Berlín: Lola Books
- 165 . Laney, D. (2001): *3D Data Management: Controlling Data Volume, Velocity, and Variety*. Nota de investigación de META Group (nº949)
- 166 . Laney, D. (2012): *The Importance of Big Data: A Definition*. Publicado por Gartner. ID del reporte: G00235055
- 167 . Larson, E. (2021): *El mito de la inteligencia artificial*. Barcelona: Shackleton Books
- 168 . LeCun, Y., Bengio, Y., & Hinton, G. (2015): *Deep Learning*. Nature, 521(7553), 436-444 <https://www.cs.toronto.edu/~hinton/absps/NatureDeepReview.pdf>
- 169 . Li, J. H. (2018): *Cyber security meets artificial intelligence: a survey*. Frontiers of Information Technology & Electronic Engineering, 19(12), 1462-1474.
- 170 . Liu, Y., & Zhang, X. (2022): *Deep Learning Algorithms for Object Detection in CCTV Footage*. Computer Vision and Image Understanding, 215, 103329.
- 171 . López de Mántaras, R. (2017): *Inteligencia artificial*. Madrid: CSIC
- 172 . López de Mántaras, R. (2023): *100 cosas que cal saber sobre intel.ligència artificial*. Barcelona: Cossetània
- 173 . Madrid Casado, C. (2020): *Filosofía de la Inteligencia Artificial*. Madrid: Digital Reasons
- 174 . Manes, F., & Niro, M. (2021): *Ser humanos: Todo lo que necesitas saber sobre el cerebro*. Barcelona: Paidós
- 175 . Manzoni, M., Medaglia, R., Tangi, L., Van Noordt, C., Vaccari, L. & Gattwinkel, D. (2022): *AI Watch Road to the adoption of Artificial Intelligence by the Public Sector: A Handbook for Policymakers, Public Administrations and Relevant Stakeholders*. Bruselas: Publications Office of the European Union. ISBN 978-92-76-52131-0, doi:10.2760/693531, JRC129100.

- 176 . Marcus, G. F. (2001): *The algebraic mind: Integrating connectionism and cognitive science*. MIT Press.
- 177 . Marcus, G. (2004). *The birth of the mind: How a tiny number of genes creates the complexities of human thought*. New York: Basic Books.
- 178 . Marcus, G. (2008). *Kluge: The haphazard construction of the human mind*. Houghton Mifflin Company.
- 179 . Marcus, G., & Freeman, J. (Eds.): (2015). *The future of the brain: Essays by the world's leading neuroscientists*. Princeton University Press.
- 180 . Marcus, G. (2024). *Taming Silicon Valley: How we can ensure that AI works for us*. Cambridge, MA: MIT Press.
- 181 . Mashey, J. (1996): *Big Data and the Next Wave of InfraStress*. Conferencia técnica de USENIX
- 182 . Maturana, H., & Varela, F. (1984). *El árbol del conocimiento: Las bases biológicas del entendimiento humano*. Barcelona: Editorial Crítica.
- 183 . Maturana, H., & Varela, F. (1980). *Autopoiesis and Cognition: The Realization of the Living*. Dordrecht: D. Reidel Publishing Company.
- 184 . Maturana, H., & Varela, F. (1973). *De máquinas y seres vivos: Una teoría sobre la organización biológica*. Santiago de Chile: Editorial Universitaria.
- 185 . Mayer-Schönberger, V., & Cukier, K. 1ed. (2013): *Big Data. La revolución de los datos masivos* Madrid: Turner
- 186 . Mayer-Schönberger, V., & Cukier, K. (2014): *Learning with Big Data: The future of education*. Houghton Mifflin Harcourt.
- 187 . Mayer-Schönberger, V., & Ramge, T. (2019): *La reinención de la economía. El capitalismo en la era del Big Data*. Madrid: Turner Noema
- 188 . Mayer-Schönberger, V., Cukier, K., & de Véricourt, F. (2021). *Framers: Human advantage in an age of technology and turmoil*. Dutton (Penguin Random House)
- 189 . McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). *A proposal for the Dartmouth summer research project on artificial intelligence*. Dartmouth College, Hanover, NH, United States.
- 190 . McCarthy, J. (1959). *Programs with common sense*. Stanford University.

- 191 . McCarthy, J., & Hayes, P. J. (1969). *Some philosophical problems from the standpoint of artificial intelligence*. Stanford University.
- 192 .
- 193 . Mearsheimer, J. (2001): *The Tragedy of Great Power Politics*. W. W. Norton & Company
- 194 . Miller, C. (2021): *La nueva guerra fría: Cómo la rivalidad entre China y Estados Unidos por la supremacía tecnológica está transformando nuestro mundo*. Madrid: Editorial Taurus
- 195 . Miller, C. (2023): *La guerra de los chips: La gran lucha por el dominio mundial*. Barcelona: Península
- 196 . Minsait. (2024). *IA, radiografía de una revolución en marcha. Ascendant: Estudio de Madurez Digital 2024*. <https://www.minsait.com/es/actualidad/eventos/informe-ascendant-2024-sobre-inteligencia-artificial-Cat>
- 197 . Minsky, M. (1961). *Steps toward artificial intelligence. Proceedings of the IRE*, 49(1), 8-30.
- 198 . Minsky, M. (1986): *La sociedad de la mente*. Buenos Aires (Argentina): Ediciones Galápagos
- 199 . Minsky, M., & Papert, S. (1988). *Perceptrons: An Introduction to Computational Geometry* (Expanded ed.). Cambridge, MA: MIT Press.
- 200 . Minsky, M. (2006): *La máquina de las emociones*. Barcelona: Debate
- 201 . Mitchell, M. (2020): *Inteligencia artificial: Una guía para pensar humanos*. Barcelona: Antoni Bosch Editor
- 202 . Mittal, S., et al. (2019): *A survey of machine and deep learning techniques for cyber security*. SN Computer Science, 1(1), 1-20.
- 203 . Moffett, S. (2007): *El enigma del cerebro*. Barcelona: Manontropo
- 204 . Moravec, H. P. (1988). *Mind children: The future of robot and human intelligence*. Cambridge, MA: Harvard University Press
- 205 . Moravec, H. P. (1993). *El hombre mecánico: El futuro de la robótica y la inteligencia humana*. Barcelona: Salvat (Obra original publicada en 1988 con el título *Mind Children: The Future of Robot and Human Intelligence*)

- 206 . More, M., & Vita-More, N. (Eds.). (2013). *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future*. Chichester, West Sussex, UK: John Wiley & Sons, Inc
- 207 . Muñoz Vela, J.M. (2021): *Cuestiones éticas de la Inteligencia Artificial y repercusiones jurídicas. De lo dispositivo a lo imperativo*. Madrid: Aranzadi Thomson Reuters
- 208 . Muñoz Vela, J.M. (2022): *Inteligencia artificial y responsabilidad penal*. Madrid: Wolters Kluwer
- 209 . Muñoz Vela, J.M. (2022): *Retos, riesgos, responsabilidad y regulación de la inteligencia artificial. Un enfoque de seguridad física, lógica, moral y jurídica*. Madrid: Aranzadi Thomson Reuters
- 210 . Muñoz Vela, J.M. (2022): *Inteligencia artificial y Cuestiones de propiedad intelectual e industrial*. Madrid: Aranzadi Thomson Reuters
- 211 . Muñoz Vela, J.M. (2022): *Avanzando en la regulación de la IA. Hacia un equilibrio entre ética, protección de los derechos fundamentales, seguridad, confianza, innovación, desarrollo y competitividad*. Madrid: Wolters Kluwer
- 212 . Muñoz Vela, J.M. (2023): *Inteligencia artificial y derecho de autor. Creaciones artificiales y su protección jurídica*. Valencia: Tirant Lo Blanch
- 213 . Muñoz Vela, J.M. (2023): *IA y responsabilidad civil. Comentarios a las propuestas europeas en materia de derechos de daños por productos defectuosos y adaptación de las normas de responsabilidad civil extracontractual*. Madrid: Aranzadi Thomson Reuters
- 214 . Muñoz Vela, J.M. (2024): *Inteligencia artificial generativa. Desafíos para la propiedad intelectual*. Madrid: UNED
- 215 . Muñoz Vela, J.M. (2024): *La regulación de la inteligencia artificial. Reto y oportunidad desde una perspectiva global e internacional*. Madrid: Aranzadi Thomson Reuters
- 216 . National Security Commission on Artificial Intelligence. (2021). *Final Report*. Disponible en: <https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf> [Informe final de la Comisión Nacional de

Seguridad sobre Inteligencia Artificial (NSCAI), presidida por Eric Schmidt como presidente y Robert Work como vicepresidente]

- 217 . Negnevitsky, M. (2011): *Artificial Intelligence: A Guide to Intelligent Systems*. Addison-Wesley
- 218 . Newell, A., & Simon, H. A. (1956). *The logic theory machine: A complex information processing system* (P-868). RAND Corporation
- 219 . Newell, A., Shaw, J. C., & Simon, H. A. (1959). *Chess-playing programs and the problem of complexity*. IBM Journal of Research and Development, 3(4), 320-335
- 220 . Newell, A., Shaw, J. C., & Simon, H. A. (1959). *Report on a general problem-solving program* (P-1584, revised). RAND Corporation.
- 221 . Newell, A., & Simon, H. A. (1972). *GPS, a program that simulates human thought*. En R. P. Banerji & M. D. Mesarovic (Eds.), *Theoretical approaches to non-numerical problem solving* (pp. 334-361). Springer
- 222 . Newell, A. (1982). *Limitations of the current stock of ideas about problem solving*. Division Eight Newsletter: Division of Personality and Social Psychology, 8(1), 2-6
- 223 . Nguyen, T. T., & Reddi, V. J. (2019): *Deep reinforcement learning for cyber security*. arXiv preprint arXiv:1906.05799.
- 224 . Nilsson, N. J. (2009): *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. Cambridge University Press
- 225 . Nonaka, I., & Takeuchi, H. (1995): *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. Oxford University Press
- 226 . Oliver, N. (2020): *Inteligencia Artificial, naturalmente: Un manual de convivencia entre humanos y máquinas*. Madrid: ONTSI
- 227 . Palma J.T. (2021): *Inteligencia Artificial*. Madrid: Ediciones Paraninfo
- 228 . Pedreño, A., Torres, A., Mora, T., & Pérez, E. del M. (2023): *1.070 Km Hub: La mayor alianza del ecosistema de innovación de España*. Publicación independiente

229. Pedreño, A., & Mora, T., et al. (2024): *La inteligencia artificial en las universidades: retos y oportunidades*. Alicante: Grupo 1 MillionBot
230. Pedreño, A., Moreno, L. (2020): *Europa frente a EE.UU. y China. Prevenir el declive en la era de la inteligencia artificial*. Amazon
231. Pedreño, A., & Moreno, L. (2023): *España en la nube: ¿Una Startup Nation o el país del desempleo juvenil?*. Amazon
232. Peguera, M. (2023): *La propuesta de Reglamento de IA: una intervención legislativa insoslayable en contexto de incertidumbre*. Editorial Reus
233. Penrose, R. (1996): *La mente nueva del emperador. En tono a la cibernética, la mente y las leyes de la física*. México: Consejo Nacional de Ciencia y Tecnología. Fondo de Cultura Económica
234. Peñarrubia, J. P. (2023): *¿Puede materializarse una ética de la inteligencia artificial, o incluso una ética digital, sin una ética profesional informática consolidada?*. XXI International Congress on Ethics and Political Philosophy / Congreso Internacional de Ética y Filosofía política: Ética, Democracia e Inteligencia artificial. Universitat Jaume I Castelló. 2-4 February 2023. Asociación Española de Ética y Filosofía Política (AEEFP)
235. Peñarrubia, Juan Pablo (2024): *¿Quién cuida el jardín?: Cuidado humanizador de los cimientos individuales y sociales ante el impacto de las tecnologías digitales*. SCIO. Revista de Filosofía, n.º 26, Julio de 2024, 123-149, ISSN: 1887-9853 DOI: https://doi.org/10.46583/scio_2024.26.1155
236. Pérez, C. (2004). *Revoluciones tecnológicas y capital financiero: La dinámica de las grandes burbujas financieras y las épocas de bonanza*. México: Siglo XXI. ISBN: 968-23-2532-3
237. Pinillos, J.L. (1969): *La mente humana*. Barcelona: Biblioteca básica Salvat.
238. Plasencia, A. (2021): *De neuronas a galaxias: ¿Es el universo un holograma?*. Valencia: Publicacions de la Universitat de València
239. Porter, M. E., & Heppelmann, J. E., (2017): *HBR's 10 Must Reads on AI, Analytics, and the New Machine Age*. Harvard Business Review Press
240. Punset, E. (2004): *Cara a cara con la vida, la mente y el universo. Conversaciones con grandes científicos de nuestro tiempo*. Barcelona: Ediciones Destino

- 241 . Punset, E. (2006): *El alma está en el cerebro. Radiografía de la máquina de pensar*. Madrid: Aguilar
- 242 . Punset, E. 1ed. (2010): *El viaje al poder de la mente. Los enigmas más fascinantes de nuestro cerebro y del mundo de las emociones*. Barcelona: Ediciones Destino
- 243 . PwC (2024): *Global Top 100 Companies - by Market Capitalisation*. <https://www.pwc.co.uk/audit/assets/pdf/global-100/global-top-100-companies-2024.pdf>
- 244 . Racionero, L. (1990): *Florecia de los Médicis*. Barcelona: Planeta
- 245 . Randstad (2024): *IA y mercado de trabajo en España*. Madrid: Randstad. Disponible en: <https://www.randstadresearch.es/ia-mercado-trabajo-espana/>
- 246 . Ratey, J. J. (2003): *El cerebro: Manual de instrucciones*. Barcelona: Mondadori
- 247 . Reglamento (UE) 2016/679 (2016): *Reglamento general de protección de datos*. <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A32016R0679>
- 248 . Restrepo D. (2018): *El camino de las redes neuronales*. Santa Marta (Colombia): Editorial Unimagdalena
- 249 . Rodriguez, A., et al. (2024): *Neural Networks for Real-Time Video Analysis in Security Applications*. Proceedings of the International Conference on Computer Vision and Pattern Recognition, 89-103.
- 250 . Rodriguez, P. (2018): *Inteligencia artificial: Cómo cambiará el mundo (y tu vida)*. Barcelona: Deusto
- 251 . Rouhiainen, L. (2018). *Inteligencia artificial: 101 cosas que debes saber hoy sobre nuestro futuro*. Barcelona: Alienta Editorial
- 252 . Rouhiainen, L. (2021). *Inteligencia artificial para los negocios. 21 casos prácticos y opiniones de expertos*. Madrid: Anaya Multimedia
- 253 . Russell, S. J. (2019): *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking

- 254 . Russell, S.J., & Norvig P. (2004): *Inteligencia Artificial: Un Enfoque Moderno* (2ª ed.). Madrid: Pearson Educación
- 255 . Russell S. J., & Norvig P. (2021): *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson
- 256 . Ruyer, R. (1984): *La cibernética y el origen de la información*. México: Fondo de Cultura Económica
- 257 . Sabouret, N. 1ed (2020). *Understanding Artificial Intelligence*. Routledge
- 258 . Salazar I. (2005): *Las profundidades de Internet: Accediendo a la información oculta*. Gijón (Asturias): Editorial Trea
- 259 . Sampedro, J. L. (2017). *El reloj, el gato y Madagascar*. Barcelona: Debate.
- 260 . Sancho, J. M., & Hernández, F. (1994). *Entrevista a Joseph Weizenbaum*. TELOS: Revista de Pensamiento, Sociedad y Tecnología, <https://telos.fundacion-telefonica.com/archivo/numero038/entrevista-a-joseph-weizenbaum/>
- 261 . Sandel, M. J. (2020). *La tiranía del mérito: ¿Qué ha sido del bien común?* Barcelona: Debate
- 262 . Schmidt, E., & Cohen, J. (2014): *El futuro digital*. Madrid: Anaya Multimedia
- 263 . Schmidt, E., & Rosenberg, J. (2015): *Cómo trabaja Google*. Madrid: Aguilar
- 264 . Schmidhuber, J. (2000). *Algorithmic Theories of Everything*. arXiv:-quant-ph/0011122. Disponible en: <https://arxiv.org/abs/quant-ph/0011122>
- 265 . Schmidhuber, J. (2015). *Deep learning in neural networks: An overview*. Neural Networks, 61, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>
- 266 . Schneider S. (2021): *Inteligencia Artificial. Una exploración filosófica sobre el futuro de la mente y la conciencia*. Badalona (Cataluña): Koan Libros
- 267 . Schwab, K. (2016): *La cuarta revolución industrial*. Barcelona: Debate
- 268 . Schwab, K. (2021): *Stakeholder Capitalism: A Global Economy that Works for Progress, People and Planet*. Wiley
- 269 . Schwab, K., & Malleret, T. (2020): COVID-19: *El Gran Reinicio*. Foro Económico Mundial
- 270 . Scherer. A. (2023): *You & AI: Hello There!. A Guide to Understanding How Artificial Intelligence Is Shaping Our Lives*. Norderstedt: Delta Labs AG Production and Publisher BoD – Books on Demand: ISBN: 978-3-37528-1361-6

- 271 . Scientific American Mind (2022, March/April): *The science of self: Neuroscientists may have discovered the brain regions that give rise to our identity. Scientific American Mind*. Disponible en: <https://www.scientificamerican.com/mind-and-brain/>"<https://www.scientificamerican.com/mind-and-brain/>
- 272 . Scientific American (2022, February): *Secrets of the mind: The science of consciousness and how our brain constructs reality. Scientific American: Special Collector's Edition*, 31(15). Disponible en: <https://www.scientificamerican.com/special-editions/>
- 273 . Scientific American (2023, June): *The brain: How this amazing organ defines our reality and helps us sense, think and act. Scientific American: Special Collector's Edition*, 32(2). Disponible en: <https://www.scientificamerican.com/special-editions/>"<https://www.scientificamerican.com/special-editions/>
- 274 . Searle, J. R. (2000). *El misterio de la conciencia*. Barcelona: Paidós.
- 275 . Seung, S. (2012): *Connectome: How the brain's wiring makes us who we are*. New York: Houghton Mifflin Harcourt.
- 276 . Sigman M. (2016): *La vida secreta de la mente: Nuestro cerebro cuando decidimos*. Barcelona: Debate
- 277 . Sigman M. (2022): *El poder de las palabras*. Barcelona: Debate
- 278 . Sigman M., & Bilinkis S. (2022): *Artificial*. Barcelona: Debate
- 279 . Silicon Valley Indicators (n.d.): *Venture Capital Investment*. <https://siliconvalleyindicators.org/data/economy/innovation-entrepreneurship/private-equity/venture-capital-investment/>
- 280 . Smith, J. (2023): *Deep Learning in Modern CCTV Systems: A Comprehensive Analysis*. *Journal of Security Technology*, 45(3), 278-295.
- 281 . Soldevilla, L. (2024): *Humanos. Las 10 cosas que la IA jamás hará por ti*. Barcelona: Profit editorial
- 282 . Sommer, R., & Paxson, V. (2010): *Outside the closed world: On using machine learning for network intrusion detection*. 2010 IEEE Symposium on Security and Privacy (pp. 305-316). IEEE.

- 283 . Stanford University Institute for Human-Centered Artificial Intelligence. (2024).
- 284 . Artificial Intelligence Index Report 2024. Disponible en <https://aiindex.stanford.edu/report/>
- 285 . State Council of the People's Republic of China (2024): *Xi urges Fujian to play pioneering role in China's modernization drive*. <https://english.www.gov.cn>
- 286 . Suleyman, M., & Bhaskar, M. (2023): *The Coming Wave: Technology, Power, and the Twenty-first Century's Greatest Dilemma*. Crown Currency
- 287 . Taddeo, M., McCutcheon, T., & Floridi, L. (2019): *Trusting artificial intelligence in cybersecurity is a double-edged sword*. *Nature Machine Intelligence*, 1(12), 557-560.
- 288 . Tangi L., van Noordt C., Combetto M., Gattwinkel D., & Pignatelli F. (2022): *AI Watch. European Landscape on the Use of Artificial Intelligence by the Public Sector*. Bruselas: Publications Office of the European Union. ISBN 978-92-76-53058-9. doi:10.2760/39336, JRC129301.
- 289 . Tascón, M., & Coullaut, A. (2016). *Big Data y el Internet de las cosas. Qué hay detrás y cómo nos va a cambiar*. Madrid: Catarata.
- 290 . Tegmark M. (2017): *Life 3.0: Being Human in the Age of Artificial Intelligence*. Knopf
- 291 . Tegmark, M. (2018): *Vida 3.0: Ser humano en la era de la inteligencia artificial*. Barcelona: Penguin Random House
- 292 . Thompson, E. (2023): *Ethical Considerations in AI-Enhanced Surveillance Systems*. *AI & Society*, 38(2), 345-360.
- 293 . Torres J. (2023): *La inteligencia artificial explicada a los humanos*. Barcelona: Plataforma Editorial
- 294 . Torrijos C. (2023): *La primavera de la Inteligencia Artificial*. Madrid: Catarata
- 295 . Turing A.M. (2012): *¿Puede pensar una máquina?*. Oviedo (Asturias): KRK Ediciones
- 296 . Varela, F. (1998). *Conocer: Las ciencias cognitivas en la vida cotidiana*. Barcelona: Gedisa.

- 297 . Varoufakis, Y. (2024): *Tecnofeudalismo: El sigiloso sucesor del capitalismo*. Barcelona: Ediciones Deusto
- 298 . Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention is all you need*. arXiv preprint arXiv:1706.03762.
- 299 . WEF: World Economic Forum (2023): *Future of Jobs Report 2023*. World Economic Forum. ISBN-13: 978-2-940631-96-4. Disponible en: <https://www.weforum.org/publications/the-future-of-jobs-report-2023/>
- 300 . Wiener, N. (1985): *Cibernética o el control y comunicación en animales y máquinas*. 2ed. Barcelona: Tusquets editores
- 301 . Wiener, N. (1988): *Cibernética y sociedad*. 3. ed. Buenos Aires: Editorial Sudamericana
- 302 . Wooldridge, M. (2009): *An Introduction to MultiAgent Systems* (2nd ed.). John Wiley & Sons
- 303 . Wooldridge, M. (2021). *A brief history of artificial intelligence: What it is, where we are, and where we are going*. Flatiron Books. ISBN: 978-1250770745
- 304 . Wooldridge. M. (2018): *Artificial Intelligence*. Michael Joseph (Penguin Random House)
- 305 . Xin, Y., et al. (2018): *Machine learning and deep learning methods for cybersecurity*. IEEE Access, 6, 35365-35381.
- 306 . Yudkowsky, E. (2018). *Map and Territory: Rationality: From AI to Zombies*, Book One. Machine Intelligence Research Institute.
- 307 . Yudkowsky, Eliezer (2023): *Pausing AI Developments Isn't Enough. We Need to Shut it All Down*. Time. <https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/>
- 308 . Yuste, R. (2023): *Lectures in neuroscience*. Columbia University Press.
- 309 . Yuste, R. (2024): *El cerebro, el teatro del mundo: Descubre cómo funciona y cómo crea nuestra realidad*. Barcelona: Paidós.
- 310 . Zuboff, S. (2020). *La era del capitalismo de la vigilancia: La lucha por un futuro humano frente a las nuevas fronteras del poder*. Barcelona: Paidós

